PERIODICAL

PROCEEDINGS

OF THE

American Philosophical Society

Held at Philadelphia for Promoting Useful Knowledge

PUBLIC LIBRARY
JUL 27 1948
DETROIT

VOLUME 92, NUMBER 3 JULY 19, 1948

CONTENTS

Migration of the Methyl Group in the Body. VINCENT DU VIGNEAUD 127

The Astronomical Origin of the Theory of Conic Sections. OTTO NEUGEBAUER 136

The Decline of Tuberculosis as the Chief Cause of Death. ESMOND R. LONG 139

Physiology of Exposures to Abnormal Concentrations of the Respiratory Gases. WALLACE O. FENN 144

Marine Molluscan Provinces of Western North America: A Critique and a New Analysis. IRWIN M. NEWELL 155

The Growth of Particles in Smokes and Clouds and the Production of Snow from Supercooled Clouds. IRVING LANGMUIR 167

The Franklin-Greene Correspondence. WILLIAM GREENE ROELKER 186

Why Union Education? Aims, History, and Philosophy of the Educational Work of the International Ladies' Garment Workers' Union. MARK STARR 194

Some Basic Issues Raised by Trade Unions. SUMNER H. SLICHTER 203

THE AMERICAN PHILOSOPHICAL SOCIETY
INDEPENDENCE SQUARE
PHILADELPHIA 6

1948

The *Proceedings* of the American Philosophical Society is published as an annual volume and ordinarily is issued in five numbers. An author desiring to submit a manuscript for publication should send it to the Editor, Luther P. Eisenhart, American Philosophical Society, 104 South Fifth Street, Philadelphia 6, Pa.

The annual subscription price is five dollars; combined with a subscription to the *Transactions* the price is ten dollars. Individual copies of the *Proceedings* are offered for sale. This issue is priced at one dollar.

LANCASTER PRESS, INC., LANCASTER, PA.

MIGRATION OF THE METHYL GROUP IN THE BODY*

VINCENT DU VIGNEAUD

Professor of Biochemistry, Cornell University Medical College

(*Read November 20, 1947*)

IN the present paper, I should like to review the experimental work which led to the inclusion of a new member in the group of indispensable nutrients, namely the simple chemical entity, the methyl group, and to present some of our recent findings regarding this grouping.

The presence of methylated compounds in the body and in foodstuffs has been known, of course, for many years, but our first inkling that methylation involved aspects not hitherto suspected occurred in some experimental work in which we were studying the conversion of homocystine to methionine in the animal body. We had produced evidence through the years that homocysteine is an intermediary step in the conversion of methionine to cystine. As our work and the work of others progressed, we had every reason to believe that demethylation of methionine to homocysteine took place in the body. The interesting question then arose as to whether this demethylation process was reversible. Could homocysteine be converted to methionine? In other words, does the body have the power not only of removing the methyl group but of adding a methyl group to the sulfhydryl grouping of homocysteine? Several investigators, White and Beach (1)[1] at Yale and later Brand (2) at Columbia, obtained results that indicated that homocystine could serve in lieu of methionine in the diet of the white rat. Through private communication we learned from Professor Rose at the University of Illinois that he likewise was obtaining similar results. When we ourselves approached the problem using a methionine-cystine free diet, we encountered the unexpected fact that the ability of homocystine to support growth in place of methionine depended upon the nature of the vitamin-B complex preparation used in the diet. On the one hand, when we supplied the B-complex entirely in a crude form, as in tikitiki, a rice bran extract, and milk vitamin concentrate, we found that homocystine could serve in lieu of methionine. On the other hand, when we furnished the B-complex mainly in the form of the purified crystalline B-vitamins then available, to minimize the amount of unknown sulfur in the diet, we found that homocystine under these conditions could not serve in place of methionine. We communicated our findings to Professor Rose and they were confirmed by Rose and Rice (3). We surmised that there must be something in the crude preparation of the vitamin-B complex which must be furnished in the diet to enable the animal to bring about the synthesis of methionine from homocysteine, and we decided to try to isolate this substance from these sources.

In our original investigation with Dyer and Kies it was noted that the livers of the rats used in the experiment with the purified vitamin supplement appeared yellowish at autopsy, and histological examination was indicative of fatty infiltration of the liver (4). Since choline was known to inhibit fatty infiltration of the liver, a suspicion arose in our minds that the diet we employed might possibly be deficient in choline. Before embarking on what promised to be a long search for the unknown factor in the vitamin-B complex preparations with Chandler and Moyer, we decided to recheck our previous results.

The previous findings were fully confirmed. The animals on the basal diet lost weight rapidly. When methionine was given, growth promptly ensued, but when homocystine was fed in place of methionine, the animals failed to grow, as we had found earlier.

In order to see what effect choline might have, it was administered to one of the animals receiving homocystine (5). We obtained a startling result; the animal which had been fed choline chloride and homocystine gained about four grams overnight and continued to grow at a rate as rapid as that of the animals which had been fed methionine. When homocystine was left out of the diet, no growth was obtained with choline.

* This paper was also presented before the Section on Chemistry Related to Food and Nutrition of the International Congress of Pure and Applied Chemistry held in London, July 21, 1947.

[1] Numbers in parentheses indicate References at end of article.

Other animals in the litter were then given choline and in every instance when choline was supplied in sufficient amounts, the animals were able to grow with homocystine in place of methionine. Further experiments with other litters amply confirmed these findings. This behavior of choline pointed toward the possibility that choline was the unknown factor for which we were about to search in tikitiki and milk vitamin concentrate. Subsequently, we were able to show that this was the case.

The effectiveness of choline in enabling homocystine or homocysteine to replace methionine in the diet and the fact that choline itself contains so many methyl groups immediately suggested to us the possibility that the animal was actually transferring methyl groups from choline to homocysteine to yield methionine. We began to grasp the fact that the body is incapable of forming this simple radical—that it must be supplied in the diet.[2]

There was a corollary to be drawn from this series of experiments. It was evident from these growth studies that animals can grow normally without being fed any choline. As long as sufficient methionine was in the diet, the animals grew at the normal rate and no deficiency symptoms were discernible, even though the diet was practically devoid of choline. It is obvious then that choline *per se* is unnecessary if sufficient methionine be present in the diet. It, therefore, seemed possible to us that choline might be synthesized in the body from methionine with the latter furnishing the methyl groups. In other words, might the transfer of a methyl group from choline to form methionine be reversible?

This hypothesis offered a reasonable explanation of the observations that had been made regarding the effect of choline, methionine, and homocystine on fatty infiltration of the liver. Best and co-workers, as you know, found that, when animals were fed a diet deficient in choline, the livers of the rats became loaded with fat. This pathological condition could be prevented by feeding choline (5*a*). Of particular interest was the fact that Tucker and Eckstein (6) at Michigan found that methionine, like choline, is capable of preventing or curing fatty liver. On the other hand, Singal and Eckstein (7) in the United States and Channon (8) in England found that homocystine, unlike methionine, does not prevent fatty liver.

This striking difference between homocystine and methionine in preventing fatty liver was, therefore, not inconsistent with the idea that the methyl group of methionine plays an important role in the synthesis of choline by the body. Thus, what was otherwise a perplexing similarity of action on the part of two apparently totally unrelated compounds such as methionine and choline was, according to this hypothesis of transmethylation, or transfer of methyl groups, readily understandable.

For experimental verification of this concept of transmethylation, we sought an experimental approach of a direct nature. It occurred to us that it might be feasible to follow the methyl group metabolically by labelling it with deuterium. In other words, if we could feed methionine with the methyl group containing deuterium in place of hydrogen, we might be able to pick up the deuteriomethyl group in the choline. We decided to synthesize deuteriomethyl iodide and methylate homocysteine with it to get the desired deuteriomethionine.

The deuteriomethyl iodide was prepared from deuteriomethyl alcohol which we synthesized by the catalytic reduction of carbon monoxide with deuterium gas. The deuteriomethyl iodide was then used to methylate the homocysteine in a liquid ammonia medium and the desired deuteriomethionine was thus formed (9).

This deuteriomethionine was fed to a white rat on a methionine-choline-free diet for three days and then the animal was sacrificed. The choline was isolated from the carcass and found to contain deuterium.

The compound was burned in a combustion tube and the water formed in the oxidation was collected and purified. The ratio of D_2O or heavy water to H_2O was then determined by the falling drop method for specific gravity. The method is very sensitive and highly accurate. One can detect two parts per million. On the

[2] It should be pointed out that an occasional animal was able to grow without choline added to the homocystine diet (5). Synthesis of choline or of some related compound by intestinal flora in such animals was considered to be the most likely explanation of the phenomenon. See also du Vigneaud, V., S. Simmonds, J. P. Chandler, and M. Cohn, Synthesis of labile methyl groups in the white rat, *Jour. Biol. Chem.* **159**: 755, 1945; Bennett, M. A., G. Medes, and G. Toennies, Growth of albino rats on a choline-free diet in which homocystine is the only sulfur-containing amino acid, *Growth* **8**: 59, 1944; Bennett, M. A., and G. Toennies, Intestinal and liver factors in the metabolic utilization of homocystine, *Jour. Biol. Chem.* **163**: 235, 1946.

basis of the methionine fed, the deuterium content of the isolated choline showed that 14 per cent of the methyl groups in the choline had been derived from the deuteriomethionine of the diet, and it looked as if our labelling of the methyl group was going to work. And, if so, this would demonstrate that the methionine could furnish the methyl groups of choline. But we had to make sure that all the deuterium was present in the methyl group before we could draw this conclusion. By degradation of the isolated choline, we were able to demonstrate that all the deuterium was present in the methyl groups of choline.

The results obtained in subsequent experiments of longer duration are given in table 1 (9). The

TABLE 1

CHOLINE ISOLATED FROM RATS FED DEUTERIOMETHIONINE

Rat No.	Time of experiment	Choline isolated	
		Atom per cent D in methyl groups	Percentage methyl groups derived from diet
	days		
325	3	11.9	13.7
226	23	50.2	57.4
326	54	59.0	67.3
387	94	74.2	88.6

deuterium content of the choline attained approximately 89 per cent of the theoretically possible value after ninety-four days.

Having established experimentally that methionine could act as a donor of methyl groups in the biological synthesis of choline, we wished to test whether choline could also serve as a methyl donor. We, therefore, synthesized deuteriocholine, and the deuteriocholine was fed to rats kept on a methionine-choline-free diet containing homocystine as the only sulfur-containing amino acid. The animals were sacrificed after several weeks on this diet. We then sought for the deuteriomethyl group in the methionine of the tissues with a successful outcome. Thus, the postulation of the synthesis of methionine from choline and homocystine *in vivo* was fully substantiated (10).

Experiments in which the labelled choline was fed with the basal diet containing ordinary methionine in place of homocystine showed that the methionine of the tissues again contained the labelled methyl group (10). Thus, we demonstrated that transmethylation took place even when sufficient methionine was supplied in the diet and when no homocystine was furnished. This experiment, taken in conjunction with other experiments in which methyl transfer occurred from methionine to choline even when ample choline was fed, established the fact that transmethylation between choline and methionine is a continual metabolic process. This continual interconversion of choline and methionine places this metabolic reaction in a type of biological phenomenon of which workers in the field of metabolism are becoming increasingly aware. Other experiments with isotopes by Schoenheimer and his co-workers have emplasized the existence of such "automatic and non-interruptable" biochemical reactions.

When we had found that choline supported the growth of animals on the homocystine diet we naturally became interested in whether the less highly methylated aminoethanol compounds, that is, the monomethyl- and the dimethylaminoethanol, could likewise support growth under these conditions. It was hoped that in the study of these particular compounds some clue might be uncovered as to the mechanism by which choline is synthesized and degraded in the course of participating in the transmethylation reaction. We found that the dimethylaminoethanol could not supply methyl groups for the synthesis of methionine from homocystine at a rate sufficiently rapid to permit as good growth of the young rat as choline did (12). However, the physical condition of the animals was obviously better, and furthermore, the dimethylaminoethanol was capable of preventing fatty livers. The monomethylaminoethanol was so toxic that we could not adequately study it.

Again to obtain a closer insight as to the fate of the methyl groups of these compounds, we turned to the labelling technique and synthesized the monomethyl- and dimethylaminoethanols with the methyl groups labelled with deuterium (12).

We found that the methyl groups of the dimethyl compound are really used very efficiently in the synthesis of choline. About 40 to 45 per cent of the methyl groups in the tissue choline were derived from the deuteriodimethylaminoethanol in three weeks, when the latter was fed as a supplement to a diet containing homocystine and cystine as the only sulfur-containing amino acids. This figure may be compared with the value of about 60 per cent obtained in experiments in which deuteriocholine was fed with homocystine (10).

That the dimethylaminoethanol prevents fatty liver and is a relatively good source of the methyl groups of choline, but is not an efficient substitute for choline in supporting the growth of young rats on a homocystine diet, would at first sight seem to be an anomalous situation. However, this situation becomes understandable if the animal is incapable of demethylating dimethylaminoethanol but can methylate it to choline.

If this should be the case, the conversion of the administered dimethylaminoethanol to choline would be accomplished on the otherwise methyl-free diet only at the expense of tissue methionine or some other source of labile methyl groups. Nevertheless, with two of the methyl groups necessary for choline synthesis already incorporated in the diet in the form of dimethylaminoethanol, the dietary situation would undoubtedly be much improved over a completely methyl-free diet. The lipotropic and the very limited growth-promoting power of dimethylaminoethanol on the homocystine diet would thus be accounted for.

On the basis of this concept, dimethylaminoethanol would not enter directly into transmethylation reactions, but would be an indirect donor of methyl groups by being convertible to choline. This concept would also carry with it the implication that choline, in engaging in transmethylation, releases only one methyl group, and in so doing it is converted to dimethylaminoethanol. Dimethylaminoethanol would thus assume a pivotal position as both the immediate precursor and the principal demethylation product of choline.

The direct conversion of dimethylaminoethanol to choline has been suggested by Jukes and Oleson (13, 14) to explain the growth-promoting and antiperotic activities of dimethylaminoethanol in chicks. These investigators suggest that dimethylaminoethanol may exist in normal diets of chicks and in other natural sources.

The results with the monomethylaminoethanol (12) may be interpreted in an analogous manner, and it appears probable that monomethylaminoethanol is also a precursor of choline but is certainly not in itself a direct methyl donor. The hypothesis that aminoethanol is transformed to choline in three successive methylation steps has been suggested by Jukes, Dornbush, and Oleson (15) on the basis of their work with *Neurospora*. This work has recently been confirmed and extended by Horowitz (16) who has presented further evidence for the validity of this mechanism of choline synthesis in *Neurospora*.

Stetten (17) has shown, through labelling with heavy nitrogen, that aminoethanol is not a product of choline degradation but is a precursor of choline. These facts likewise fit in with our concept of the demethylation of—and the resynthesis of—choline. The difference in behavior of the partially methylated aminoethanols and choline, in so far as the ease with which an intact methyl group may be transferred, may be ascribed to the fact that choline is a quaternary nitrogen compound, while the others are trivalent nitrogen compounds.

In our first work on the support of growth of rats on the homocystine diet with choline, we of course turned to investigate betaine (18), since we knew that betaine occurred in nature. We found that it was capable of supporting growth on the homocystine diet, although not quite as efficiently as choline. Recently we have studied betaine labelled both in the methyl groups with deuterium and in the glycine moiety with heavy nitrogen, N^{15} (19). The isotopic betaine was fed to two growing rats at the level of fifty-five milligrams daily for two weeks on an otherwise methyl-free diet. At the end of the experimental period, more than 40 per cent of the methyl groups of the tissue choline had been derived from the betaine fed. The betaine was almost as effective a source of methyl groups of tissue choline as dietary choline itself. On the other hand, only .5 per cent of the nitrogen of the choline had been derived from the dietary betaine. In fact, the N^{15} concentration of the glycine of the tissue protein was twice as high as that of the choline. It is obvious that the betaine was not directly converted to choline. This is in agreement with the data of Stetten (20) who showed that the nitrogen of betaine finds its way to choline by way of glycine and aminoethanol.

Comparable experiments with dimethylglycine demonstrated that transmethylation from this compound to choline occurred to only a very small degree (19). The same was likewise found with the monomethylglycine, that is, sarcosine (21).

When we first recognized the existence of a metabolic interrelationship between choline and methionine we suggested that "the methionine-homocystine-choline relationship may be significant, not only to sulfur and fat metabolism, but also to the other problems of methylation in the animal organism" (5). We regarded the formation of creatine as one of the most important and obvious of these methylations. In the experi-

ments, therefore, in which we had fed methionine containing the isotopic methyl group we saw the opportunity of testing crucially whether methionine was linked with creatine metabolism by isolating the creatine from the tissues and seeing whether it contained the deuteriomethyl group (9). The results we obtained are shown in table 2. These experiments offer clean-cut proof that the methyl group of methionine is utilized in the synthesis of creatine.

TABLE 2

CREATINE ISOLATED FROM TISSUES OF RATS FED DEUTERIOMETHIONINE

Rat No.	Time of experiment	Creatine isolated	
		Atom per cent D in methyl groups	Percentage methyl groups derived from diet
	days		
325	3	7.9	9.1
326	54	56.7	64.8
387	94	73.0	87.1

Evidence for the role of methionine in creatine synthesis has also come from an independent line of attack by Borsook and Dubnoff in tissue slice experiments. They found that additional creatine was synthesized from guanidoacetic acid when methionine was added to the medium (22). In more recent work, they have demonstrated that this methylation is dependent on oxygen and is inhibited by oxidation inhibitors (23). It is extremely interesting that they have found that choline and betaine are ineffective as direct methyl donors in this methylation (22). However, they have since shown that extra creatine is formed if, in addition to choline or betaine, homocystine or homocysteine is added to the liver slices (23). This suggests that choline can supply methyl groups for creatine synthesis indirectly by an intermediate formation of methionine. This hypothesis is strengthened by their recent demonstration that methionine is formed from homocysteine, homocystine or homocysteine thiolactone by homogenates of rat liver if either choline or betaine is added as a source of methyl groups (24). This methylation is independent of oxygen.

In the intact animal we would expect that the methyl group of choline should eventually find its way to creatine on the basis of the experimentally established transfer of methyl groups from choline to methionine. This expectation was fulfilled in an experiment where we fed deuteriocholine (9). Examination of the creatine of the tissues and the creatinine in the urine disclosed the presence of the isotopic methyl group.

With the demonstration of the source of the methyl group of creatine all the precursors of this important compound are now known, a solution of a problem that has been attacked by many workers during the past several decades. It has been clearly shown by the work of Bloch and Schoenheimer (25) and by the work of Borsook and Dubnoff (26) that the amidine moiety of arginine is transferred from arginine to glycine to form guanidoacetic acid. The former workers used the tracer method with the stable nitrogen isotope and the latter group the tissue slice technique.

The metabolic relationships among the methyl compounds which I have so far discussed may be schematically represented in the manner shown in figure 1. As can be seen, the methyl group of methionine can be released for methylation, such as the methylation of guanidoacetic acid to creatine, and as a result homocysteine is formed. The homocysteine may then accept a methyl group from choline yielding methionine which in turn can give up its methyl group. Thus, the methyl groups of choline are available for creatine synthesis. The irreversibility of the methylation of guanidoacetic acid, which we were able to demonstate, is represented. We fed creatine containing the deuteriomethyl group and found no deuterium in the choline of the tissues (27). Undoubtedly, the cleavage of the methyl group from methionine is a far more complicated affair than indicated here. There is some evidence to indicate that ATP—adenosine triphosphate—is involved (28). The intimate details of the removal of the methyl group and the enzyme systems involved are being actively worked on in many laboratories and we can expect much information in the next few years on this question.

From the evidence I have already discussed, we believe that the choline yields one methyl group leaving dimethylaminoethanol which must be remethylated to choline again to enter into the actual process of transmethylation (12). Dimethylaminoethanol acts as precursor and degradation product.

The dimethylaminoethanol would come from the step-wise methylation of aminoethanol. Stetten has shown by the use of serine (28*a*) and of glycine (17) containing isotopic nitrogen that aminoethanol is derivable from these compounds.

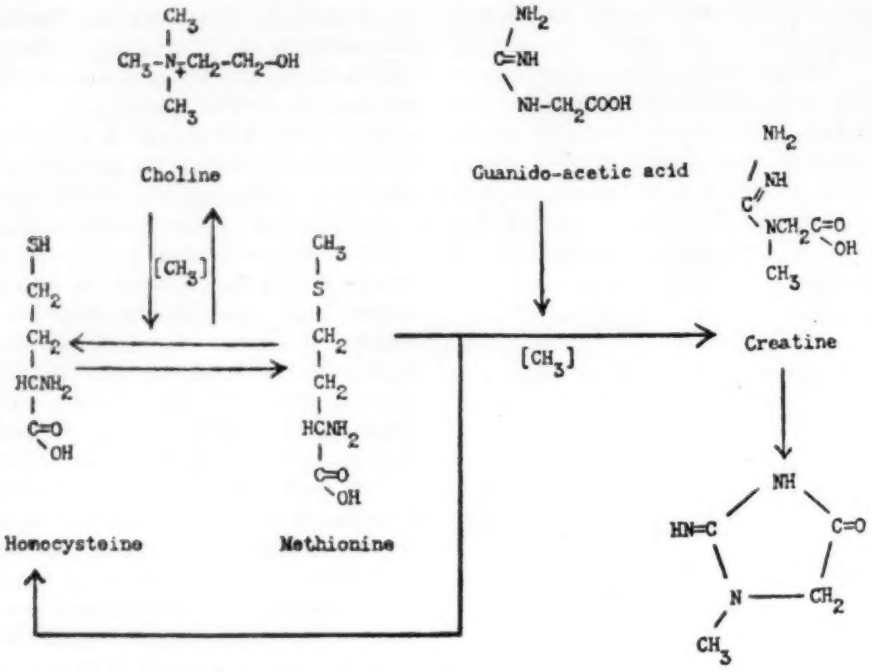

FIG. 1. Scheme of the metabolic interrelationships of methionine, choline, and creatine.

The data of Shemin on serine labelled with C^{13} in the carboxyl group and N^{15} demonstrated that glycine may arise from serine through a loss of the β-carbon of serine (29).

The interrelationship of methionine and choline, which has given a rational basis for the understanding of why methionine should have a lipotropic action similar to choline, likewise affords an explanation of their similar qualitative behavior in preventing hemorrhagic kidneys. This condition develops in rats kept on a choline-deficient diet. This pathological deficiency syndrome was studied in detail by Griffith and co-workers (30) and later it was the subject of an interesting investigation by Welch (31). The lability of the methyl group of methionine and choline has been called upon by Griffith and Mulford in their suggestion that a dietary supply of labile methyl groups is nec. sary to prevent hemorrhagic kidneys (32). But the specificity of structure in regard to lipotropic and antihemorrhagic activity is not nearly so pronounced, as it is in the case of methyl donors.

A compound may be effective as a lipotropic or antihemorrhagic agent without regard to lability of the methyl or even the presence of a methyl group, but any compound which is a methyl donor is effective by making possible the synthesis of choline. Although methyl donors are not unique in preventing fatty livers and hemorrhagic kidneys, the deficiency symptoms are indicative of a lack of methyl donors ordinarily present in the diet. These substances as far as we know are choline, betaine, and methionine, and possibly dimethylaminoethanol.

Whether the metabolic reactions involved in transmethylation occur only in the rat or whether they proceed in other species as well is a question of considerable significance. It has been shown by Klose and Almquist that, in the chick, choline plus homocystine will substitute for methionine in the diet (33). But it has been found by Jukes that choline is required even though methionine be present in the diet of chicks in the prevention of perosis, the slipped tendon syndrome, produced by choline deficiency (34). Aminoethanol plus methionine will not replace choline in the diet of the chick (34*a*), but monomethyl- or dimethylaminoethanol will replace choline if methionine is present (14). Jukes suggests that the chick cannot methylate aminoethanol to monomethylaminoethanol but can methylate the latter compound in two steps to give choline using methionine as the source of methyl groups. In extending the investigation of transmethylation to other species, we found that in the rabbit, as in the rat, trans-

methylation reactions proceed. When deuteriomethionine was fed to a rabbit, deuterium was found in the isolated choline and creatine of the tissues and creatinine of the urine (34*b*).

That the process of transmethylation likewise occurs metabolically in man has been established in our laboratory. This is of particular significance since considerable interest has been aroused concerning the therapeutic implications of choline and methionine in man, particularly in pathological conditions of the liver. The possible application of the results of transmethylation studies on the rat to nutritional pathological problems in man has been predicated on the tacit assumption that similar metabolic routes are followed in the human being.

Because of the growing importance of this possibility, we felt that a direct attempt ought to be made to test whether or not the process of transmethylation really takes place in normal man. We have made a few exploratory experiments in this direction, applying the technique of labelling the methyl group to the human, and have found that the process of transmethylation does, in fact, take place in man (35). We have fed deuteriomethionine to a man for three days and then collected three hundred and fifty cubic centimeters of blood from him and isolated the choline from his blood. The urine was collected daily and the creatinine was isolated from the urine. Both choline and creatinine showed the presence of the deuteriomethyl group, far beyond experimental error, thus demonstrating that transmethylation is a metabolic process in man. The work also demonstrated that man can synthesize choline.

It is obvious from the number and variety of methyl compounds that we are only at the beginning of unraveling the metabolic interrelations of the methyl compounds that are involved in the metabolism of the animal body, and we believe many can be studied. Quantitatively we believe the evidence indicates that it would be possible that the methionine, choline, and betaine of the diet could furnish the methyl groups for synthesis of the other methyl compounds in the body, since many of them occur in traces. Our hypothesis at the moment is that these compounds are related metabolically with methionine and choline and are dependent on this source of methyl groups in the diet for their synthesis. The possibilities involved are intriguing, particularly when one thinks of such relationships as with epinephrine, and also in connection with certain detoxication reactions involving methylation.

Our initial opportunity to test the hypothesis that other N-methyl compounds of the body derived their methyl groups from the methionine of the diet came in our transmethylation studies in the rabbit (34*b*). With this animal we had the opportunity of isolating another of the N-methyl compounds which occur in the body, namely anserine, which contains a methyl group attached to the imidazole ring of the histidine moiety of carnosine. It is a constituent of muscle tissue occurring in fairly large amounts.

After twenty-eight days approximately 2 per cent of the methyl groups of the total anserine had been derived from the dietary deuteriomethionine, a value far beyond the experimental error. The conversion was much less than in creatine, which was 20 per cent, indicating a slower rate of turnover. Thus, we have evidence that this compound is related in metabolism to methionine and in turn to choline. And another field of metabolism is linked to those we have already mentioned. The results give greater credence to our hypothesis that all the N-methyl and S-methyl compounds of the body tissues have their source of methyl groups in the methionine and choline and betaine of the diet.

Finally, I should like to tell you of some very recent experiments which have put us in position to detect the fate of the methyl groups in the diet that are not excreted in the urine in the form of creatinine. We were anxious to find out whether the labile methyl groups could be oxidized to carbon dioxide. This we have been able to do by labelling the methyl group with radioactive carbon. Melville, Rachele, and Keller (36) in my laboratory worked out procedures for converting CO_2 to CH_3I and thus we were able to make methionine and other methylated compounds that we desired.

The C^{14} was obtained from the Oak Ridge laboratories in the form of carbonate which was converted to bicarbonate and reduced to formate by hydrogen under pressure in the presence of palladium. The formate was then converted to the methyl formate and reduced to methyl alcohol by hydrogen with a copper-chromite catalyst. Methyl alcohol was then converted to methyl iodide by hydriodic acid, and the methyl iodide thus formed passed into a liquid ammonia solution containing the sodium homocysteinate. The yields throughout these reactions were highly satisfactory.

Recently we carried out, in collaboration with Mackenzie, Chandler, Keller, Rachele, Cross, and Melville (37), an experiment in which we administered two hundred milligrams of this radiomethyl L-methionine to a 165-gram rat on an otherwise normal diet and then collected the expired CO_2 for fifty-two hours. We started out using this large dose because we wanted to make sure that we could detect any CO_2 coming from the methyl group of the methionine. We found that approximately one-third of the methionine methyl group was oxidized to CO_2 during the period of collection.

With the radiomethyl we were also in position to determine a complete balance of the radiocarbon of the methyl group fed with the radiocarbon in the various organs and excreta. The results were eminently satisfactory. We were able to account for 98 to 99 per cent of the administered radiocarbon. Radioactivity determinations on the separate tissues and organs of the rat indicated that the methyl carbon had been most highly concentrated in the kidneys, liver, and adrenals in the order named, while intermediate concentrations were found in the intestines, pancreas, and testes.

The radiomethyl also provides a more effective tool for determining whether those N- and S-methyl compounds found in the body in only minute amounts derive their methyl groups from the methionine and choline of the diet. One such compound is epinephrine, and we have obtained, with Keller, some preliminary data to show that the methyl group of methionine can be utilized in the formation of epinephrine.

We hope to go forward with the determination of the concentration of the radiomethyl in the various methyl compounds of the body. We also hope to ascertain if the methyl group is used in other synthetic reactions of the body. It is also obvious that we have the opportunity of studying other methylated compounds as for example, radiomethyl creatine and radiomethyl choline.

In closing I think it would be fair to state that the process of transmethylation as a metabolic process has been established through these researches beyond a peradventure of doubt. Methylation in the body has been known for many years, but I would like to emphasize that the concept of a requirement for methyl groups in the diet was a new concept in nutrition. This concept was based on the recognition that the body is incapable of synthesizing a methyl group for the purpose of certain methylations and that an adequate supply of preformed methyl groups in a utilizable or "biologically labile" form was required in the diet for this purpose.

The unique position of the methyl group as a dietary essential is more fully appreciated when it is realized that its significance is not restricted to only one field of metabolism. This vital grouping plays a role in nitrogen metabolism, in sulfur metabolism, in fat metabolism, and in carbohydrate metabolism. Thus, it can be seen that the process of transmethylation—or migration of methyl groups—truly transects all of these metabolic pathways.

REFERENCES

1. White, A., and E. F. Beach. 1937. The role of cystine, methionine, and homocystine in the nutrition of the rat. *Jour. Biol. Chem.* **122**: 219.
2. Brand, E. 1938. Growth response to sulfur amino acids. *Jour. Biol. Chem.* **123**: xv.
3. Rose, W. C., and E. E. Rice. 1939. The utilization of certain sulfur-containing compounds for growth purposes. *Jour. Biol. Chem.* **130**: 305.
4. du Vigneaud, V., H. M. Dyer, and M. W. Kies. 1939. A relationship between the nature of the vitamin B complex supplement and the ability of homocystine to replace methionine in the diet. *Jour. Biol. Chem.* **130**: 325.
5. du Vigneaud, V., J. P. Chandler, A. W. Moyer, and D. M. Keppel. 1939. The ability of homocystine plus choline to support growth of the white rat on a methionine-free diet. *Jour. Biol. Chem.* **128**: cviii.

   du Vigneaud, V., J. P. Chandler, A. W. Moyer, and D. M. Keppel. 1939. The effect of choline on the ability of homocystine to replace methionine in the diet. *Jour. Biol. Chem.* **131**: 57.

5*a*. Best, C. H., and C. C. Lucas. 1943. Choline-chemistry and significance as a dietary factor. *Vitamins and Hormones* I: 1. N. Y., Academic Press.

6. Tucker, H. F., and H. C. Eckstein. 1937. The effect of supplementary methionine and cystine on the production of fatty livers by diet. *Jour. Biol. Chem.* **121**: 479.
7. Singal, S. A., and H. C. Eckstein. 1939. Supplementary proteins and amino acids and dietary production of fatty livers in mice. *Proc. Soc. Exp. Biol. and Med.* **41**: 512.
8. Channon, H. J., M. C. Manifold, and A. P. Platt. 1938. The effect of sulfur-containing amino-acids on liver fat deposition. *Chem. and Ind.* **57**: 600.
9. du Vigneaud, V., J. P. Chandler, M. Cohn, and G. B. Brown. 1940. The transfer of the methyl group from methionine to choline and creatine. *Jour. Biol. Chem.* **134**: 787.

   du Vigneaud, V., M. Cohn, J. P. Chandler, J. R. Schenk, and S. Simmonds. 1941. The utilization of the methyl group of methionine in the biological synthesis of choline and creatine. *Jour. Biol. Chem.* **140**: 625.

10. SIMMONDS, S., M. COHN, J. P. CHANDLER, and V. DU VIGNEAUD. 1943. The utilization of the methyl groups of choline in the biological synthesis of methionine. *Jour. Biol. Chem.* **149**: 519.
11. SCHOENHEIMER, R. 1942. The dynamic state of body constituents. Cambridge, Harvard Univ. Press.
12. DU VIGNEAUD, V., J. P. CHANDLER, S. SIMMONDS, A. W. MOYER, and M. COHN. 1946. The role of dimethyl- and monomethylaminoethanol in transmethylation reactions in vivo. *Jour. Biol. Chem.* **164**: 603.
13. JUKES, T. H., and J. J. OLESON. 1945. Choline-like nutritional effect of dimethylaminoethanol. *Jour. Biol. Chem.* **157**: 419.
14. JUKES, T. H., J. J. OLESON, and A. C. DORNBUSH. 1945. Observations on monomethylaminoethanol and dimethylaminoethanol in the diet of chicks. *Jour. Nutrition* **30**: 219.
15. JUKES, T. H., A. C. DORNBUSH, and J. J. OLESON. 1945. Some observations on nutritional effects of choline and related compounds. *Federation Proc.* **4**: 157.
16. HOROWITZ, N. H. 1946. The isolation and identification of a natural precursor of choline. *Jour. Biol. Chem.* **162**: 413.
17. STETTEN, D., JR. 1941. Biological relationships of choline, ethanolamine, and related compounds. *Jour. Biol. Chem.* **138**: 437; **140**: 143.
18. CHANDLER, J. P., and V. DU VIGNEAUD. 1940. The comparative action of choline and betaine in effecting the replacement of methionine by homocystine in the diet. *Jour. Biol. Chem.* **135**: 223.
19. DU VIGNEAUD, V., S. SIMMONDS, J. P. CHANDLER, and M. COHN. 1946. A further investigation of the role of betaine in transmethylation reactions in vivo. *Jour. Biol. Chem.* **165**: 639.
20. STETTEN, D., JR. 1941. Biological relationships of choline, ethanolamine, and related compounds. *Jour. Biol. Chem.* **140**: 143.
21. DU VIGNEAUD, V., S. SIMMONDS, and M. COHN. 1946. A further investigation of the ability of sarcosine to serve as a labile methyl donor. *Jour. Biol. Chem.* **166**: 47.
22. BORSOOK, H., and J. W. DUBNOFF. 1940. The formation of creatine from glycocyamine in the liver. *Jour. Biol. Chem.* **132**: 559.
23. BORSOOK, H., and J. W. DUBNOFF. 1945. Methylation of guanidoacetic acid by homocystine plus choline with rat liver slices. *Jour. Biol. Chem.* **160**: 635.
24. BORSOOK, H., and J. W. DUBNOFF. 1947. Methionine formation by transmethylation in vitro. *Jour. Biol. Chem.* **169**: 247.
25. BLOCH, K., and R. SCHOENHEIMER. 1941. The biological precursors of creatine. *Jour. Biol. Chem.* **138**: 167.
26. BORSOOK, H., and J. W. DUBNOFF. 1941. The formation of glycocyamine in animal tissues. *Jour. Biol. Chem.* **138**: 389.
27. SIMMONDS, S., and V. DU VIGNEAUD. 1945. A further study of the lability of the methyl group of creatine. *Proc. Soc. Exp. Biol. and Med.* **59**: 293.
28. BINKLEY, F. 1947. Phosphorylation of methionine. Presented before the American Chemical Society at New York. September 16.
28*a*. STETTEN, D., JR. 1942. The fate of dietary serine in the body of the rat. Jour. Biol. Chem. **144**: 501.
29. SHEMIN, D. 1946. The biological conversion of L-serine to glycine. *Jour. Biol. Chem.* **162**: 297.
30. GRIFFITH, W. H. 1941. The relation of choline to the kidneys. Biological Symposia V: 193. Lancaster, Jaques Cattell Press.
31. WELCH, A. D., and R. L. LANDAU. 1942. The arsenic analogue of choline as a component of lecithin in rats fed arsenocholine chloride. *Jour. Biol. Chem.* **144**: 581.
32. GRIFFITH, W. H., and D. J. MULFORD. 1940. Choline, creatine and the "labile methyl" supply. *Proc. Soc. Exp. Biol. and Med.* **45**: 657.
GRIFFITH, W. H., and D. J. MULFORD. 1941. Choline metabolism. VI. Hemorrhagic degeneration and the labile methyl supply. *Jour. Amer. Chem. Soc.* **63**: 929.
32*a*. MOYER, A. W., and V. DU VIGNEAUD. 1942. The structural specificity of choline and betaine in transmethylation. *Jour. Biol. Chem.* **143**: 373.
33. KLOSE, A. A., and H. J. ALMQUIST. 1941. Methionine in the diet of the chick. *Jour. Biol. Chem.* **138**: 467.
34. JUKES, T. H. 1940. Effect of choline and other supplements on perosis. *Jour. Nutrition* **20**: 445.
34*a*. ——. 1941. The effect of certain organic compounds and other dietary supplements on perosis. *Jour. Nutrition* **22**: 315.
34*b*. SCHENCK, J. R., S. SIMMONDS, M. COHN, C. M. STEVENS, and V. DU VIGNEAUD. 1943. The relation of transmethylation to anserine. *Jour. Biol. Chem.* **149**: 355.
35. SIMMONDS, S., and V. DU VIGNEAUD. 1942. Transmethylation as a metabolic process in man. *Jour. Biol. Chem.* **146**: 685.
36. MELVILLE, D. B., J. R. RACHELE, and E. B. KELLER. 1947. A synthesis of methionine containing radiocarbon in the methyl group. *Jour. Biol. Chem.* **169**: 419.
37. MACKENZIE, C. G., J. P. CHANDLER, E. B. KELLER, J. R. RACHELE, N. CROSS, D. B. MELVILLE, and V. DU VIGNEAUD. 1947. The demonstration of the oxidation in vivo of the methyl group of methionine. *Jour. Biol. Chem.* **169**: 757.

THE ASTRONOMICAL ORIGIN OF THE THEORY OF CONIC SECTIONS

OTTO NEUGEBAUER

Professor of the History of Mathematics, Brown University

(*Read November 20, 1947*)

INTRODUCTION AND SUMMARY

APPOLONIUS's theory of the conic sections (about 220 B.C.) is undoubtedly one of the masterpieces of ancient mathematics and will remain one of the great classics of mathematical literature. Very little, however, is known about the origin of the theory of conic sections as such. It is well known that the familiar names of these curves, ellipse, hyperbola, and parabola, originated from Apollonius's method of attack, which consists in applying the methods of "geometrical algebra" to the discussion of these curves. Apollonius obtains his curves by intersecting a fixed skew circular cone by a plane of variable angle. We also know that this approach is very different from the earliest known method to obtain conic sections. Menaechmus, a pupil of Eudoxus, is credited with the discovery of the conic sections (about 350 B.C.). These curves were obtained, however, by a very peculiar construction. The cone is a right circular cone; the intersecting plane is always perpendicular to one of the generating lines of the cone, and the three types of curves are obtained by varying the angle at the vertex of the cone.

The strange condition of perpendicularity of the intersecting plane always seemed to me to point to only one explanation, the theory of sundials. The generating line must be the "gnomon," the intersecting plane is the plane on which the shadow is cast. I shall demonstrate in the following section that a very simple type of sundial satisfies all requirements which lead to the above definition of conic sections.

THEORY OF THE SUNDIAL

In this section I shall develop the consequences of the following hypothesis. A gnomon of length 1 is adjusted in such a way that it always points at the sun when it culminates. The plane onto which the shadow is cast is perpendicular to the gnomon. We want to find the length of the shadow as function of time.

Let us first say that this arrangement has obvious advantages. In the case of the equinoxes the gnomon lies in the plane of the equator. The receiving plane always intersects the plane of the horizon exactly in the East-West line. The shadow at equinoxes travels a straight line from West to East. If α is the angle which the sun is distant from noon, then the length s of the shadow is simply $\tan \alpha$ and can be found graphically without any theory (*cf.* fig. 1). If

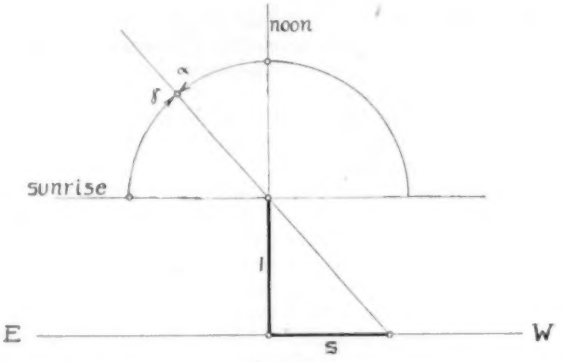

FIG. 1.

we measure time from sunrise, an angle $90 - \alpha = \gamma$ corresponds to the moment α before noon. Thus $s = \cot \gamma$. In the case of a solar declination different from zero the above description is no longer accurate. The deviation, however, remains obviously small because the adjustment of the gnomon towards the culminating point of the sun keeps the deviation of the shadow surface from the plane of the daily solar orbit within narrow limits. Hence we may expect that $s = \cot \gamma = \tan \alpha$ requires only small corrections. We furthermore see that the receiving plane needs only to be of small dimensions perpendicular to the main West-East path of the shadow. Finally the adjustment towards the culminating point is very easy to control; one must merely prevent the noon shadow from becoming visibly different from zero. Thus the whole construction is very simple in practical execution.

For our purposes, it is essential, however, to develop the accurate theory for arbitrary declination δ between $+\epsilon$ and $-\epsilon$, ϵ being the obliquity

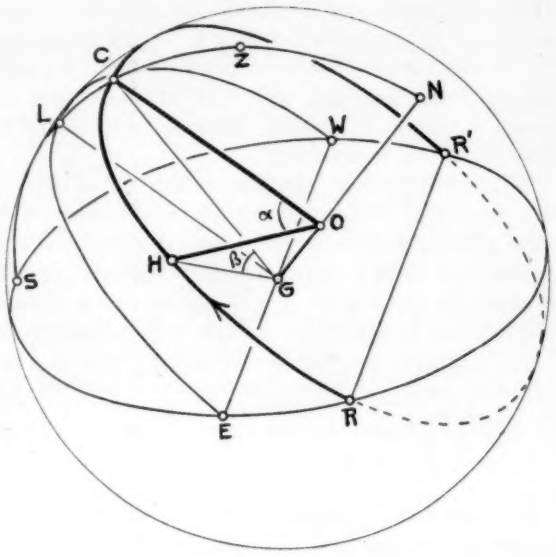

FIG. 2.

of the ecliptic. Figure 2 represents the celestial sphere with the shadow-casting point G of the gnomon as center. ELW is the equator, RESWR′ the horizon, RHCR′ the daily orbit of the sun, culminating at C. Hence CGL = δ the declination, and CG the direction of the gnomon. Assume the sun in H. The angle α, measuring the distance from noon is given by α = COH, where O is the center of the parallel circle RHCR′. The length s of the shadow is then given by $s = \tan\beta$, where β = HGC is the angle between the ray HG and the direction of the gnomon GC. All we have to do is to find β as function of α.

In figure 3 the parallel circle RHCR′ is once more represented—though, for the sake of simplicity, in horizontal position. A line perpendicular to its plane at the center O must meet the point of the gnomon G. Hence CG is the extended gnomon. From figure 2 it follows that OCG = CGL = δ.

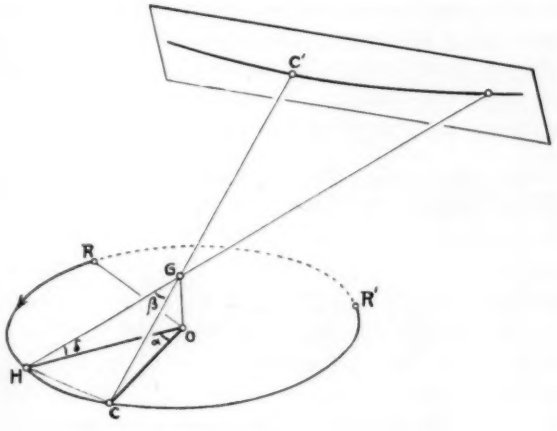

FIG. 3.

It is now easy to express the angle β = HGC as function of α = COH. If r = OH is the radius of the parallel circle we have on the one hand (fig. 3)

$$\mathrm{CH} = 2r \sin\frac{\alpha}{2}.$$

Because $\mathrm{CG} = \mathrm{HG} = \dfrac{r}{\cos\delta}$, we have on the other hand

$$\mathrm{CH} = \frac{2r}{\cos\delta}\sin\frac{\beta}{2}.$$

Thus $\sin\dfrac{\beta}{2} = \cos\delta \cdot \sin\dfrac{\alpha}{2}$. Because $s = \tan\beta$ we compute also $\tan\dfrac{\beta}{2}$. By a simple computation one finds

$$(1) \qquad \tan\frac{\beta}{2} = \frac{\cos\delta \sin\dfrac{\alpha}{2}}{\sqrt{1 - \cos^2\delta \sin^2\dfrac{\alpha}{2}}}.$$

This answers our question.

CONSEQUENCES

(A) Because $\cos\delta = \cos(-\delta)$ we obtain the same shadow length for declinations symmetric to the equator, especially for $\delta = \epsilon$ and $\delta = -\epsilon$. Our sundial shows equal shadow lengths for both solstices.

(B) Because the geographical latitude has no influence on δ, our sundial gives the same shadow lengths for all localities on the earth.

(C) Because shadows are only cast on the receiving plane when $|\beta| < 90°$ we find for $|\delta| = \epsilon$ the following limit for α. We substitute in formula (1) the value $\tan\dfrac{\beta}{2} = 1$ and find

$$(2) \qquad \cos\alpha_0 = 1 - \frac{1}{\cos^2\epsilon}.$$

Because $1/\cos^2\epsilon > 1$ we see that $\cos\alpha_0 < 0$ and therefore $\alpha_0 > 90°$.

The angle c = ROC measures the half length of daylight for a solar declination δ. If $c > \alpha_0$ the sundial does not operate for the interval

from $\alpha = c$ to $\alpha = \alpha_0$. The farther we go north the longer is c for the summer solstice and the later after sunrise the sundial can be used. For latitudes in the eastern basin of the Mediterranean, however, this is of little interest.

(D) Because δ is small, thus $\cos \delta$ close to 1, we see from (1) that $\tan \frac{\beta}{2}$ is not very different from $\tan \frac{\alpha}{2}$. This confirms our expectation that the length of the shadow is in all cases close to $\tan \alpha$.

(E) Example. For $\epsilon = 23;50$ we obtain the following table for the shadows at summer solstice.

α	β	s
0°	0°	0
20	18;17	.33
40	36;28	.74
60	54;26	1.40
70	65; 3	2.15
80	72; 1	3.08
90	80;36	6.04
100	88;58	55.44
108	95;28	

The limiting angle is $\alpha_0 = 101°$ ($\beta = 90°$). The greatest possible value for α at Babylon is about 108°. Hence the sundial is in practice always usable. At the winter solstice, α reaches only the value of about 72°. At that time the shadow at sunrise begins with a length only slightly more than twice the length of the gnomon. At equinox the shadow at sunrise is 6. Thus we see that our sundial has the additional advantage of avoiding very long morning shadows for at least half of the year.

THE CONIC SECTIONS

It follows from figure 3 that the shadow-casting rays HG form the generating lines of a straight circular cone of slant δ. The shadow cone is the other sheet of this cone with the vertex G. The shadow of G falls onto a plane perpendicular to the generating line which continues CG; this continuation GC′ is the gnomon. Thus the shadow of G travels on a conic section one point of which is the foot C′ of the gnomon. The form of the conic section depends on the slant δ or on the angle $180 - 2\delta$ at the vertex G, which is the only variable parameter for the discussion of the shadow curve.

This corresponds exactly to the definition of the conic sections given by Menaechmus. The only generalization beyond our description lies in the fact that δ is no longer restricted to the interval form $-\epsilon$ to $+\epsilon$ if we want to obtain more than hyperbolae. This generalization is trivial as soon as one investigates from a geometrical point of view the curves travelled by the shadow.

CONCLUSION

Though I feel confident that the above explanation gives the real motivation for the early Greek theory of conic sections, I must admit that I do not know of the existence of sundials of this type. The majority of Greek sundials with a plane as receiving surface assume either a horizontal or a vertical position of this plane. The only exception is one sundial in the British Museum whose receiving plane coincides with the equator plane,[1] while the gnomon points towards the North or South pole of the celestial sphere. The shadow travels on circles for all declinations different from zero, whose radius is the same for opposite signs of δ, again independent of the geographical latitude. This sundial does not work, however, for the equinoxes and, in practice, for a considerable interval before and after the equinoxes. One might say that this rather unpractical clock represents the other extreme of a development whose opposing end is indicated by the arrangement discussed in the preceding sections. The common basis of the theory of both clocks is an arrangement as given in figure 2 though interchanging the role of gnomon and receiving plane.

[1] *Cf.* Joseph Drecker, Theorie der Sonnenuhren, chap. V, Berlin, De Gruyter, 1925.

THE DECLINE OF TUBERCULOSIS AS THE CHIEF CAUSE OF DEATH

ESMOND R. LONG

Director, The Henry Phipps Institute, University of Pennsylvania
Director of Research and Therapy, National Tuberculosis Association

(*Read April 23, 1948*)

THE decline of tuberculosis as the chief cause of death has been a major, but little appreciated, event in human history. The causes of mortality, in general, have been of little historical interest. The written record of man's long social evolution largely ignores the means that remove most of the population from the earth. Death is seldom of more than individual and local concern when it results from what are called natural causes. Only when it is violent does it excite general interest.

Disease, also, enters the usual historical record only when it occurs in dramatic form. Then, however, it is of absorbing interest. Few professional histories of the world omit mention of the plague of Athens or the Black Death. Both were devastating catastrophes terrifying at the time and responsible for profound social change. But in the grand total of extinction of human life they were only passing episodes.

This comment is not meant to minimize the destruction of lives by violent means or epidemic disease. Wars, the uncontrollable disasters caused by the natural elements, and the frequent eruptions of malignant plagues have swept untold millions of people from the earth. But quietly, steadily and almost unnoticed, during the same length of time tuberculosis has killed billions.

No Thucydides or Daniel Defoe has yet come forward to highlight tuberculosis. A less widely known but reliable authority, however, has made a calculation of its total mortality that is worth remembering. Robert Koch, one of the founders of the science of bacteriology, in the classic paper of 1882 announcing the discovery of the tubercle bacillus, recorded his estimate that one-seventh of all people died of tuberculosis. Great strides have been made in the control of the disease since 1882, and the tuberculosis mortality rate in progressive countries is far below what it was then, but in the vast Orient it is still responsible for this staggering proportion of all deaths.

TUBERCULOSIS FROM ANCIENT TIMES TO THE INDUSTRIAL REVOLUTION

The origins of tuberculosis are as difficult to trace as the beginnings of human life. We only know that in man's first compact populations it was already endemic. Ancient Hindu accounts mention it, as do the records from Mesopotamia and Egypt. More tangible evidence exists in the presence of still recognizable tuberculous residues in the bones of Egyptian mummies. Unfortunately for medical history the internal organs could not be preserved. Since tuberculosis is principally a disease of the lungs, and in our civilization is at least forty times as frequent there as in the bones, the tuberculous lesions still visible in the skeletal remains of high-class Egyptians suggest a wide prevalence of pulmonary tuberculosis in the general population.

In ancient Greece and Rome, where celebrated physicians left imperishable records, tuberculosis was a common disease. Although Hippocrates and Galen and their contemporaries left no statistics on its prevalence, modern understanding of the spread of tuberculosis, a contagious disease, makes it certain that it must have been disseminated extensively under the prevailing living conditions.

The written record is scanty for the long period from the fall of Rome to the age of printing. But expert conjecture again fills in some of the gap. In a remarkable essay on the history of tuberculosis the late Allen Krause made what appear highly plausible estimates of the rise and fall of tuberculosis in the migrations, urbanization and changing economy of Western Europe. With the disruption of the Western Roman Empire in the fifth and sixth centuries decentralization of population occurred, with migration from the cities and development of separate, widely dispersed agricultural economies. Communities were small and communications poor. The conditions for dissemination of the disease were lacking and presumably the prevalence declined. With the crafts-

man industrialization of the Middle Ages, however, centralization of population recurred, with intense crowding in small walled towns. Probably tuberculosis increased rapidly and reached high endemic rates.

A time came finally when municipal governments kept a record of their annual deaths. With a good deal of uncertainty in detail but fair general reliability, these can be translated into modern terms. It is customary today to measure the mortality from tuberculosis in deaths per year per hundred thousand population. From the fragmentary records available mortality curves for different countries at different times have been constructed. There is reason to believe, according to Krause, that the maximum mortality occurred in England about 1780, when the death rate for "consumption" was 1,120 per hundred thousand population. This high rate (about twenty times the current rate in England) was probably related to the beginning of the industrial revolution. This revolution has occurred in different countries at different times. Russia is in an early phase of it now. In general the industrial revolution has led first to crowding, excessive competition, and lowered standards of living, and later to labor organization for protection, development of civic responsibilities, and a rise in the standard of living. Coincidently tuberculosis has risen and then declined. It appears to have decreased almost continuously in England since the latter years of the eighteenth century.

TUBERCULOSIS IN PRIMITIVE PEOPLES

Long after Europe was heavily seeded with tuberculosis, migrations to America, Africa, the Pacific, and the Arctic spread it in an astonishing series of epidemics. Decimating waves carried off North and South American Indians, and would have been even more destructive had not smallpox and other acute diseases killed more rapidly, and had the communications between native tribes been easier. Lest it be thought that the role of tuberculosis is here exaggerated, a relatively recent Indian epidemic, well studied and recorded, may be cited. In 1880 when buffalo hunting suddenly ceased in the northwest, some thousands of Indians from the great plains of Canada, in order to prevent their starvation, were herded into stockades on the Qu'Appelle reservation in Saskatchewan. Tuberculosis, a sporadic disease in the days of plains migration, became epidemic immediately from contagion in close quarters. Within two years the death rate had reached the appalling rate of nine thousand per hundred thousand population per year, or thirty times the white rate at that time in Canada and three hundred times the present white rate in the United States. For a few years nearly ten per cent of the Qu'Appelle Indian population died of tuberculosis each year. Slowly the rate subsided, reaching 270 in 1932, fifty years after the peak.

In Polynesia, tuberculosis, brought in with the crews of sailing ships, was a stunning catastrophe. A well informed observer, who had the disease himself, Robert Louis Stevenson, described, in an essay that has all the value of a scientific document, the virtual obliteration of island populations. Like others he stressed the severity of the disease in physically powerful people.

There is no need to elaborate specifically on this interesting phase in the history of tuberculosis. It is important, however, to understand the significance of these epidemics, witnessed in our own time, for their bearing on the past history of the disease. Tuberculosis in our white population today is still serious, but it is far less formidable in killing power than when it first struck the Eskimos and South Sea Islanders, or the Senegalese soldiers imported from Africa in World War I, whose susceptibility excited much comment in medical journals. The extraordinary malignancy of tuberculosis in primitive peoples meeting tuberculosis for the first time furnishes a clue to the past history of all peoples. No person in the world is immune to tuberculosis. No two people in the world are exactly alike in their resistance to it. The stage was set long ago for the vast operation of natural selection in the contest between tuberculosis and man. We have reason to believe that the racial groups of today stand in different stages of the survival of the fittest. Indians, Negroes, Eskimos, and Pacific Islanders are in mid-stages. The white race, after five thousand years of association with the tubercle bacillus, is in a late stage. Long since in all races, however, mortality from tuberculosis has been much more influenced by environmental than biological factors. Those environmental factors will be discussed from here on. For simplification their effects will be illustrated chiefly by their operation in the United States.

DECLINE OF TUBERCULOSIS IN THE UNITED STATES AND THE MODERN WESTERN WORLD

In our country tuberculosis was a highly destructive disease a century ago, and the principal cause of mortality. It caused from a sixth to a fifth of all deaths (Morton). It is responsible for less than a twenty-fifth today, and is seventh instead of first in the list of causes of death. The middle years of the century witnessed a dilution of the population by young immigrants from Europe, chiefly of rural stock, who migrated to the farms in this country. The dilution reduced the average tuberculosis death rate (Krause), but the rate rose again when the rural population flocked to the cities with the great intensification of industry in the 1880's. Crowding and poor hygiene at first favored widespread infection. Later, as in England after the first phases of the industrial revolution, living standards rose and tuberculosis again declined. Figure 1 illustrates the course of tuberculosis mortality in three large American cities from 1812 to 1938. Figure 2 shows the mortality rate for the United States as a whole from 1900 to 1946.

In the last fifty years the mortality rate from tuberculosis has dropped more than eighty per cent in the United States—from two hundred to less than forty per hundred thousand population. The reasons for this great decrease have been much debated. During those years an increasingly effective organized campaign has been waged against the disease, originating in the formation of the National Tuberculosis Association in 1904, a large voluntary body of physicians and laymen, with a present contributed annual income of about nineteen million dollars, devoted to education of the public on tuberculosis, and stimulation of official governmental programs against the disease. In those fifty years there has been a steady increase in the establishment of official tuberculosis control units in city, county, and state health departments, and in 1944 a tuberculosis control division was set up in the United States Public Health Service, with the avowed objective of hastening the campaign for the eradication of the disease by assistance to states and promotion of research.

Most important of all has been the provision of hospitals and sanatoria for the care of tuberculosis. More than a hundred thousand beds are available for treatment and isolation, as compared with less than ten thousand at the opening of the century. The voluntary quarantine effected through the use of these beds by patients with open disease has been of incalculable benefit in suppressing contagion.

In recent years the campaign to supply adequate hospitalization has been supplemented by active case-finding programs in hundreds of communities. Prior to the last few years the overwhelming majority of cases were discovered in late stages when cure was difficult or impossible to achieve. The present campaign, based on refined diagnostic procedures, including the tuberculin test and particularly the small, inexpensive x-ray films now used in extensive mass surveys, has resulted in the discovery of many thousands of cases in the earliest, symptomless, and most readily curable stage.

Simultaneously the treatment of the disease has improved greatly, with introduction of a variety of mechanical and surgical procedures for immobilizing an affected lung and putting it at rest, collapsing it entirely or removing it in part or whole (pneumothorax, thoracoplasty, lobectomy,

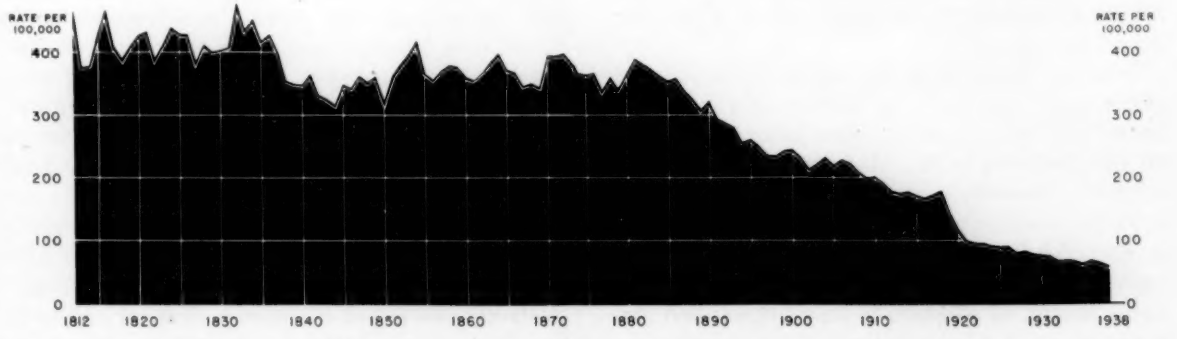

FIG. 1. Pulmonary Tuberculosis death rate 1812–1938, New York (Manhattan and Bronx Boroughs only), Philadelphia, and Boston. Original data compiled by the Prudential Insurance Company of America for period 1812 to 1912, thereafter by Research Service, New York Tuberculosis and Health Association, Inc.

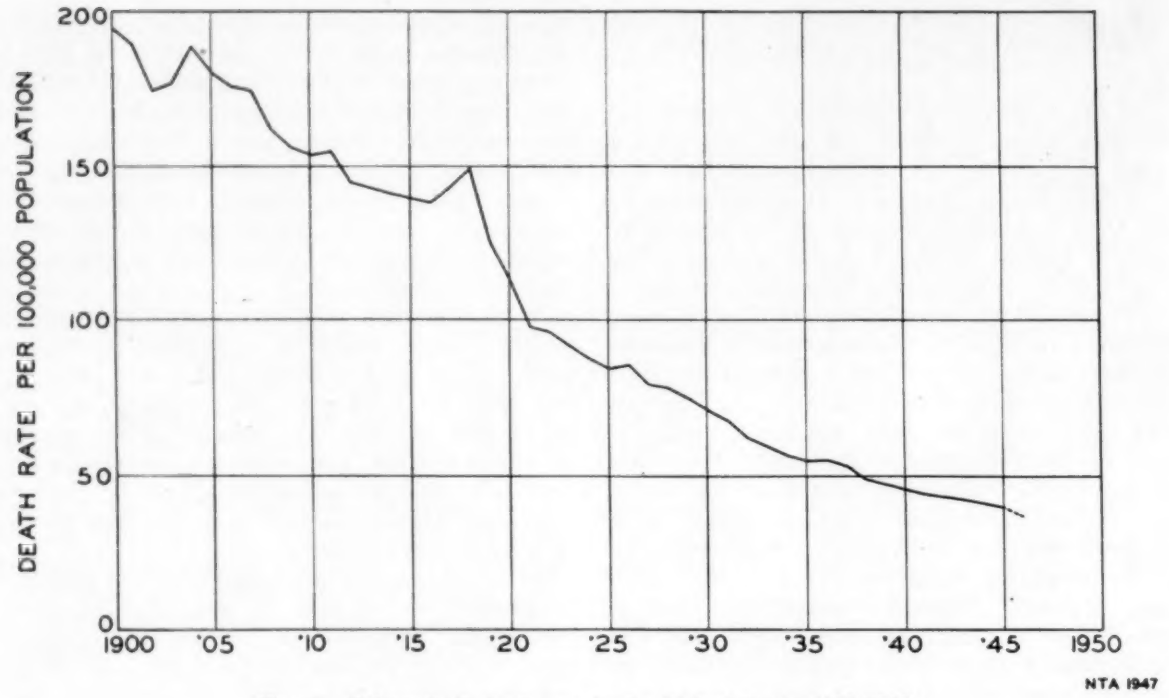

FIG. 2. Tuberculosis death rate in the United States 1900–1946.

and pneumonectomy). The most encouraging advance of recent years has been the introduction of the antibiotic drug, streptomycin, which has been shown to be remarkably successful against certain forms of the disease.

Of the effectiveness of this specific campaign against tuberculosis there can be no doubt. No one, however, believes it solely responsible for the great decline in tuberculosis. The steady rise in the standard of living, including such factors as improved nutrition for the population as a whole, better housing for a significant part of the population, less exposure to the elements, and greatly improved general hygiene and sanitation, has been of fully as great importance as specific anti-tuberculosis measures.

Prior to World War II, and to a considerable extent since the war, the pattern of tuberculosis control outlined above has been followed throughout the western world. During the war neutral countries proceeded with their normal program. In lesser measure occupied countries, not ravished by war, did likewise. An unusually strong program is in operation in the Scandinavian countries. There, in addition to case-finding and generous provision of hospitalization for isolation and care, a vigorous campaign of vaccination is under way, in which an attenuated form of the tubercle bacillus is used (BCG, so named after its discoverers, Calmette and Guérin in France), which appears to confer a significant degree of protection. This method is widely practiced in South America also. The Danish Red Cross plans to extend its vaccination program to the war-torn countries of Europe. In the United States, where the other methods described have been strikingly successful, vaccination thus far has been practiced on a small scale only.

Throughout western Europe energetic case-finding programs are defining the problem as never before. This is true both of the former neutrals and the nations recently at war, victors and vanquished alike. Most countries report an increase in known cases of disease. The British attribute the rise to better work in bringing cases to light; the Germans are gratified at their own success in this respect, but attribute much of the increase to war-time conditions. Time will tell which is the more important factor.

War and prolonged devastation almost invariably raise tuberculosis rates. Conspicuous rises occurred in World War I. Greater increases, relatively, occurred in World War II, although the actual levels attained were not so high in the second conflict as the first, when the rise took place from an originally much higher plane. In

devastated Warsaw, the rate trebled between 1939 and 1944, reaching five hundred per hundred thousand. In Vienna the recorded rate rose from 109 in 1938 to 257 in 1945, and in Berlin from 77 to 264. Bad as conditions are, however, beginning recovery is already evident. The war-time rise in England and France was checked in 1941. Even in badly bombed Germany, desperately in need of housing, in war-torn Italy and in most of eastern Europe rates have declined since 1945.

The speed at which the mortality drops with restoration of peace is so startling as to require special explanation. The seeds sown by the deprivations of war may be expected to mature for some time. Yet death rates have fallen sharply. Holland was saved from starvation in the very nick of time in 1945, and that year had twice as high a death rate from tuberculosis as before the war. Yet today it has one of the lowest tuberculosis death rates in the world. The explanation is that the appalling conditions of 1945 killed off many persons with chronic tuberculosis who would have died under normal conditions in the subsequent three or four years. There is much to support the view that this has happened repeatedly elsewhere.

THE FUTURE OF TUBERCULOSIS

The present favorable position in the United States needs no further comment here except to say that we are still far from the position where we can slacken in our effort. With the knowledge at hand, and that expected, tuberculosis ought to become a rare disease here within a few more decades, instead of killing fifty thousand persons a year as at present. It will not die out abruptly, however, for present success in therapy is now preserving the lives of many chronically ill persons not likely ever fully to recover, whose names will appear in the mortality tables for tuberculosis as late as the year 2000,—i.e., unless chemotherapeutic or other research discovers much more effective means of treatment of the disease than we have at present. This, indeed, may happen.

The future of tuberculosis in Europe is not predictable. If there were political and economic stability permitting the operation of public health programs, the death rate would drop steadily. Tuberculosis as a population problem is so sensitive, however, to adverse circumstances that predictions in the light of the present disturbed world affairs are impossible. The fantastic currency inflation which occurred in Germany in 1922–1923, and reduced thousands to poverty, illustrates the danger. The death rate, which had declined since the war-time peak in 1918, rose sharply in 1922. It dropped again with financial restabilization and resumption of normal employment.

If reasonable stability is maintained in the western world, however, the present understanding of tuberculosis as a preventable contagious disease and the demonstrated effectiveness of control methods should not only remove it from the top of the list of causes of death, as happened in the United States forty years ago, but push it far down to a minor position in mortality.

For the rest of the world, where standards of living are still abysmally low, substantial improvement cannot be expected until social and economic advance occurs, ensuring more nearly adequate nutrition and housing, revolutionizing hygiene and sanitation, and permitting the wide application of specific measures for tuberculosis control.

REFERENCES

KRAUSE, A. K. 1928. Tuberculosis and public health. *Amer. Rev. Tbc.* **18**: 271–322.

MORTON, S. G. 1834. Illustrations of pulmonary tuberculosis. Phila., Key and Biddle.

PHYSIOLOGY OF EXPOSURES TO ABNORMAL CONCENTRATIONS OF THE RESPIRATORY GASES[1]

WALLACE O. FENN

Professor of Physiology, University of Rochester School of Medicine and Dentistry

(*Read November 20, 1947*)

I HAVE had the temerity to select the whole broad field of the physiology of the respiratory gases for my subject but obviously there are only certain limited aspects with which I shall be able to deal. My particular purpose is to make use of a new type of diagram for the illustration and analysis of certain problems of respiration. This particular method of plotting data has proved itself exceedingly useful in our laboratory for the understanding of the rather complicated mathematical process of the regulation of the alveolar air of the lungs by the breathing. I shall try to show you how useful this chart is and to point out the reasons why it is better for this purpose than are other more familiar methods of plotting. Incidentally I may be able to describe to you in terms of this chart the results of some of our various experiments in the field of respiration.

The chart in question represents carbon dioxide and oxygen pressures in the alveolar air as ordinates and abscissae respectively. Such a chart, as shown in figure 1, gives a bird's eye view of the field of respiratory gases. Any combination of concentrations of these gases can be represented by a suitable point on the chart, and any respiratory experiment involving changes in the concentrations of these gases can be represented by a line. Man is adapted to function normally over a rather small range as indicated in the center of the chart. Outside of this central region is another area within which consciousness is generally retained but not without serious deterioration of the performance, especially near the borders. Anywhere outside of this second area consciousness is quickly lost and convulsions or death occur regardless of the cause. This might indicate that in all cases the final cause of loss of consciousness was the same—possibly lack of oxygen in the brain. This might occur either because of lack of oxygen in the air or because of a defective circulation to the brain as a secondary result.

The dangers of high CO_2 might occur by breathing too slowly, by obstruction of the trachea or drowning or by breathing in a closed space. If the CO_2 tension becomes higher than 10 per cent consciousness is lost in ten minutes or less, for CO_2 has a narcotic effect in high concentrations (Seevers, 1944). On the contrary too much ventilation of the lungs leads to acapnia or lack of CO_2. The symptoms of acapnia are tingling in the fingers, toes, or the whole body, twitching in the face, muscular incoordination, a feeling of lightheadedness, and finally muscular spasms or tetany.

On the summit of Mt. Everest or in an airplane at a sufficient altitude the lack of oxygen results in collapse from anoxia. This can hardly occur from rebreathing in a closed space unless the volume of that space is as small as about 100 liters because in larger volumes the carbon dioxide reaches toxic levels before the oxygen can be dangerously reduced. In small volumes sufficient CO_2 can be retained by the alkaline reserve of the body so that oxygen depletion can proceed to lethal limits.

While we die quickly without sufficient oxygen, the chart shows that we can also have too much of a good thing because in deep diving operations oxygen poisoning is a familiar hazard and accounts for many cases of erratic behavior of men working under water at great depths. In 1938 I had the thrill of reading Commander Ellsberg's account of his underwater enterprises entitled *Men Under the Sea*. I was somewhat skeptical when I read (p. 124) that in the high oxygen pressure of the compressed air which the diver required on the bottom "the fat of his body is literally burned out of him after a few deep dives." While this expression presumably was not intended to be a scientific statement of fact it nevertheless set me to wondering what the effect would

[1] Much of the work referred to in this paper was supported in part by a contract between the University of Rochester and the Air Materiel Command at Wright Field. It is a pleasure also to acknowledge my indebtedness to my colleagues Dr. Hermann Rahn and Dr. Arthur B. Otis who have contributed largely to the material of this lecture.

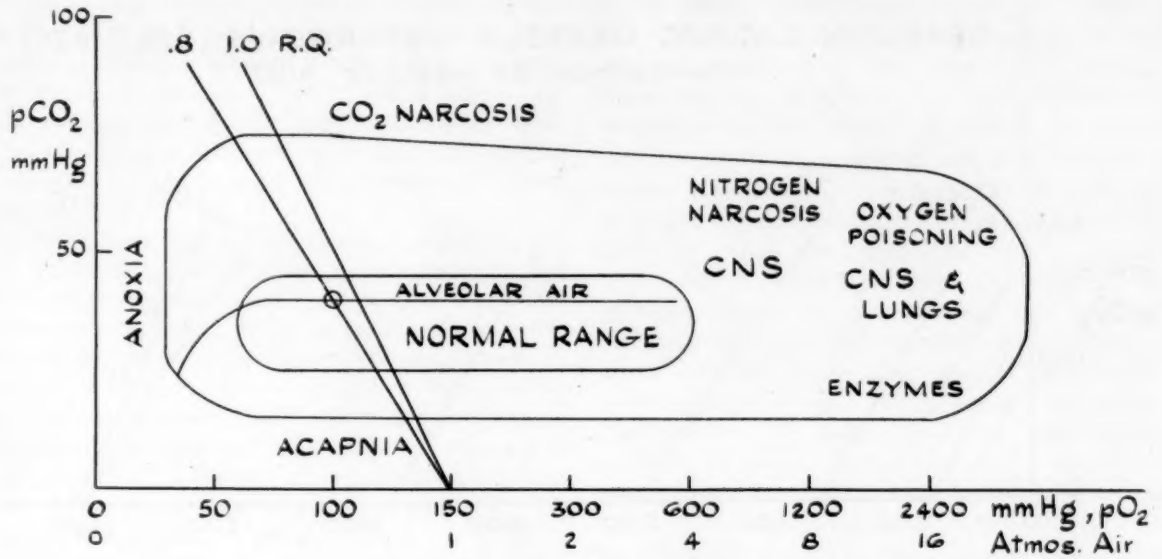

FIG. 1. Diagram to illustrate approximately the normal and abnormal ranges of CO_2 and O_2 concentrations in the alveolar air in the lungs.

really be on the rate of oxygen consumption of tissue if the oxygen pressure were increased. The natural inference of course is that the rate of burning would be increased but experiment showed that just the reverse was actually true.

Oxygen consumption is usually measured either by pressure or volume in a small glass vessel or respirometer but this technique is difficult at high pressures. We accordingly planned to measure the rate of CO_2 output as an indication of metabolic rate. This was done by absorbing the carbon dioxide in a solution of barium hydrate and measuring its electrical conductivity. This apparatus, containing a bit of muscle or kidney or skin of a frog, was enclosed in a heavy bomb, built to stand 50 to 60 atmospheres of pressure. With the help of Miss Elizabeth Cass, now of Russell Sage College, a series of measurements was made which showed that, after a brief initial increase, the CO_2 output was diminished when the oxygen pressure was increased. The higher the pressure the lower the final level. When the oxygen pressure was returned to normal the rate of metabolism of the tissue increased again. The same pressure change caused by nitrogen did not have this effect. Before these results were published (Cass, 1947), the same finding was reported by Dr. Stadie (1945) of this city. The chain of reactions involved in the oxidation of foodstuffs evidently functions properly only at certain intermediate pressures of oxygen. Whether this inhibition of the enzymes which occurs at high oxygen pressures causes the central nervous system symptoms and the edema of the lungs which seems to be the immediate cause of death in oxygen poisoning, is a debatable question because we do not know whether these effects occur *in vivo* soon enough and at a low enough pressure to explain the symptoms.

Nitrogen poisoning is a less well defined entity, but a definite deterioration in the mental activity and performance of divers has been reported by Behnke (1935) and by J. B. S. Haldane (with Case, 1941) and their colleagues. The reports demonstrate convincingly that we are dealing with a real phenomenon. It occurs at oxygen pressures too low to cause oxygen poisoning. It is not due to pressure *per se* because it does not occur in a similar pressure of helium. The result is generally interpreted in terms of the lipoid solubility of the nitrogen. When the pressure is sufficient to force the necessary number of nitrogen molecules into solution in the lipoid layers of the brain cells, narcosis appears as with any other narcotic in spite of the chemical inertness of this molecule. During the past year Dr. Charlotte Haywood has been looking for this phenomenon of nitrogen narcosis in simpler systems such as white blood corpuscles phagocytizing starch grains, and isolated nerves responding to electrical stimulation but we have unfortunately not yet

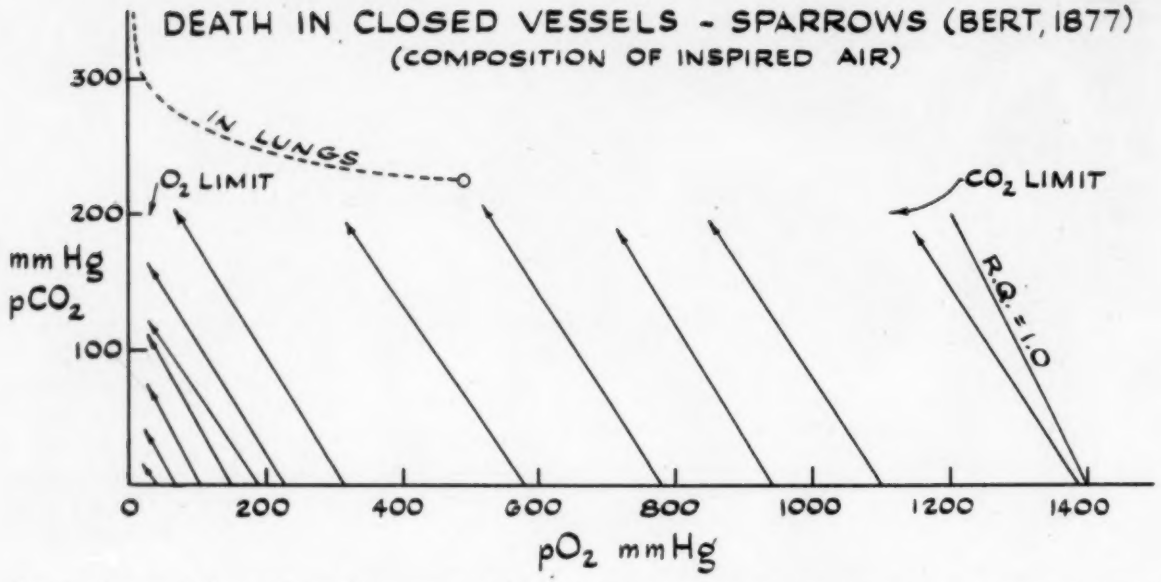

FIG. 2. Arrows indicate changes in pressure of CO_2 and O_2 in the gas in a closed vessel in which a sparrow is confined. The tip of the arrow marks the death of the sparrow.

been able to establish any effect of the nitrogen with certainty. Close observations of the behavior of the microscopic crustacean *Daphnia* when exposed to these pressures of 800 lbs. of nitrogen did not reveal any definite effects. These experiments are continuing and we have hopes now of finding some positive effects but the recent experiments are too preliminary to warrant their mention at this time.

It is noteworthy that the effects of high CO_2 are additive to the effects of both high and low O_2 as indicated by the rounded corners of the areas outlined in figure 1. The effects of low CO_2 are also additive to the effects of low O_2; probably low CO_2 effects are additive to those of high O_2 but this is purely guesswork at the present time.

Through the area on the chart which represents the normal range there runs the alveolar air line which represents the normal level of carbon dioxide in the air deep inside the lungs as a function of barometric pressure or altitude. The normal alveolar point, marked by a circle, lies on a diagonal coming from a point on the base line which represents the composition of the inspired air at ground level with $pO_2 = 149$ and pCO_2 practically zero. If, in the lungs, oxygen is removed from this air and carbon dioxide is added in exactly equal amounts its composition will change along the line marked R.Q. = 1.0.[2] In the normal metabolism of the body, however, the CO_2 output is only 80 or 85 per cent of the oxygen consumed. We can imagine therefore that as the inspired air is taken into the lungs it is "processed" by loss of O_2 and gain of CO_2 in such proportions that its composition changes along the line marked R.Q. = 0.8 until it reaches the alveolar point. Actually when we inspire the alveolar air is slightly diluted so that the alveolar point moves slightly down the 0.8 diagonal and then returns to the normal value before the next inspiration.

The alveolar air line is nearer the top of the normal range than the bottom and this is significant of one dominant fact in the control of respiration. When the CO_2 goes slightly above the alveolar air line it stimulates an overpowering urge to breathe. Breathing dilutes the alveolar air and lowers the pCO_2 to the normal value. On the other hand if one grows excited and breathes abnormally deeply and frequently, as some aviators have done under combat conditions, the alveolar point moves down the diagonal toward the region of acapnia without however provoking any inhibitory sensation. The control of respiration

[2] R.Q. or sometimes Q is the abbreviation for respiratory quotient, this being the ratio of the CO_2 given out to the O_2 taken in.

is therefore a balance between an unconscious desire to breathe as little as possible and a definite stimulus which is aroused when CO_2 is too high or the O_2 is too low. The stimulus due to low O_2 is responsible for the way in which the alveolar air line dips down as the oxygen tension falls below 60 mm Hg or as the altitude increases above 18,000 feet.

The oxygen-carbon dioxide diagram serves well to illustrate some pioneer experiments of the French physiologist Paul Bert published in 1878. These are described in his famous book on *Barometric Pressure* in the chapter entitled "Death in Closed Vessels." The experiments were simple and involved placing sparrows under bell jars at different pressures and with different amounts of oxygen added to the air. After closing the chamber the sparrow used oxygen and gave off CO_2 in accordance with the metabolic R.Q. (or CO_2/O_2 ratio) until eventually the carbon dioxide concentration became so high or the oxygen concentration so low that death ensued. When the animal died a sample of the air was taken and analyzed. In the chart (fig. 2) each diagonal line represents the changing composition of the air produced by one bird, starting at the base line and proceeding upwards to the left to the point of the arrow where breathing stopped. Thereafter the air in the bell jar suffered no further change but the air in the lungs rapidly lost its remaining oxygen and gained CO_2 (as indicated by the dotted line drawn for one of the cases) until the heart also stopped and death ensued a minute or too later. It can readily be seen that at initial oxygen pressures, below 200 mm, the bird died when the oxygen pressure of the inspired air diminished to about 25 mm Hg. At pressures above 200 mm on the other hand the carbon dioxide reached a lethal concentration of about 200 mm before the oxygen could be depleted to lethal levels. Why sparrows can tolerate 200 mm of CO_2 while man loses consciousness at 70 or 80 mm in ten minutes or less we cannot say. Note that in the sparrow as in man the slope of these lines is less than the slope of 1.0 at which the loss of O_2 and the gain of CO_2 are equal.

The survival period in a closed vessel naturally depends upon the size of the vessel. If a man were sealed up hermetically in the Grand Central Station it would require about twenty years before the CO_2 level reached a value of 10 per cent (if I have guessed the size of the station correctly).

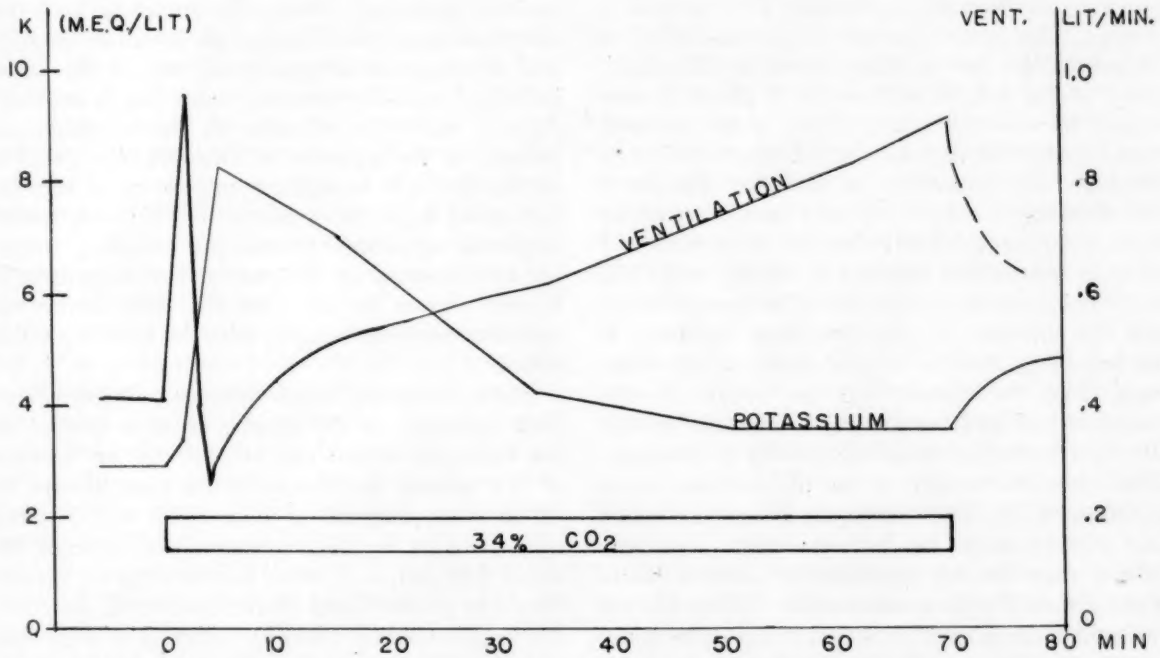

FIG. 3. Effects of inhaling 34 per cent CO_2 in air on anesthetized cats. Abscissae: time in minutes after the beginning of the CO_2 period. Ordinates: concentration of potassium in the blood plasma in m.eq. per liter (left) and pulmonary ventilation in lit. per minute (right).

If this concentration were reached quickly it would cause unconsciousness in about ten minutes. With twenty years for adaptation a man might tolerate a far higher concentration. In our laboratory Dr. Otis and Mr. Chapin spent three days in a gas tight room in an atmosphere of 3 per cent CO_2 and developed a definite decrease in their sensitivity to carbon dioxide as judged by their rate of ventilation and alveolar carbon dioxide tension. Similarly two dogs were kept in this room for two weeks in a concentration of carbon dioxide which rose gradually to 11 per cent. In this time the adaptation which they showed was not sufficient to permit any further increase in the level of CO_2. They lost all appetite for food and were definitely morbid but recovered promptly when they were restored to normal conditions. A considerably longer time would presumably be required for significant further adaptation. Seevers (1944) however has raised rats in increasing concentrations of CO_2 until some of them were able to withstand 20–25 per cent for twenty-nine days but not without the loss of half the body weight.

Acute exposure to very high carbon dioxide tensions of 30 or 40 per cent provides good surgical anesthesia for some animals like rats but in man there is said to be too much tendency to cause convulsions. Recovery from such CO_2 narcosis is prompt. An acute exposure to concentrations of this order has far reaching effects on the physiology of the animal and results in death if continued for several hours. The blood pressure rises rapidly and then falls and a heart block may develop. The breathing is also first stimulated and then depressed. This may be illustrated by some experiments of Miss Joanna Mackay (1947) on anesthetized cats exposed to 34 per cent CO_2, and illustrated in figure 3. The graph shows that in addition to the transitory increase in the breathing there is a sudden escape of potassium from the tissues into the blood. In the course of half an hour this potassium level gradually falls while the breathing gradually increases. Meanwhile the acidity of the blood increases to a pH of 6.78. The blood sugar and the lactic acid concentrations are both elevated. It is surprising that life can continue for several hours when the acidity is so abnormal. When CO_2 is removed there is another small rise in potassium, presumably due to changes in membrane equilibrium between the inside and outside of the cells. The potassium liberated by CO_2 probably comes from the liver and is absent in an eviscerated animal. Since the pattern of the breathing in an eviscerated animal is the same as in the intact animal the depression of breathing which occurs cannot be ascribed to the potassium.

In the experiments of Paul Bert with closed vessels the birds died not directly because the inspired air had become lethal but because with this particular inspired air the sparrow was no longer able to control the composition of the alveolar air within non-lethal limits by increasing the respiration. It is the alveolar air which is most important physiologically because it is this air which comes into equilibrium with the blood. I should like therefore to direct your attention now to various ways in which the alveolar point can be made to move on the oxygen-carbon dioxide diagram.

The first case is that of holding the breath which has been well worked out by my colleagues, Drs. Rahn and Otis (1947). In a general way we may regard this experiment as similar to that of survival in a closed vessel with the difference that the vessel in this case is reduced to the size of the lung. This situation is illustrated in the next chart (fig. 4) plotted on the same coordinates as before. The alveolar air line represents the starting point at which breath holding begins at various altitudes. When the breath is held the composition of this alveolar air increases in CO_2 and decreases in oxygen along one of the paths indicated until the breaking point line is reached. At this point the stimulus to the breathing, as defined by the equation of Gray (1946) (shown on the chart) is enough to cause a rate of ventilation about eight times normal. This very useful empirical equation expresses the breathing stimulus as a function of CO_2 excess and oxygen lack. It was shown by Dr. Otis that this theoretical equation coincides roughly with the breaking point curve.

Three different breath holding pathways have been indicated on the chart. In pure oxygen at the right the breath can be held for an average of 153 seconds and the pathways must always lie on an R.Q. diagonal of slope equal to 1.0 if the total pressure remains constant, even though the actual CO_2 output is much less than the O_2 intake. In air at ground level in the middle of the chart the breath holding pathway indicates a large loss of oxygen for very little increment in carbon dioxide, the average R.Q. being only 0.1 and the average time 43 seconds. The low R.Q. is due

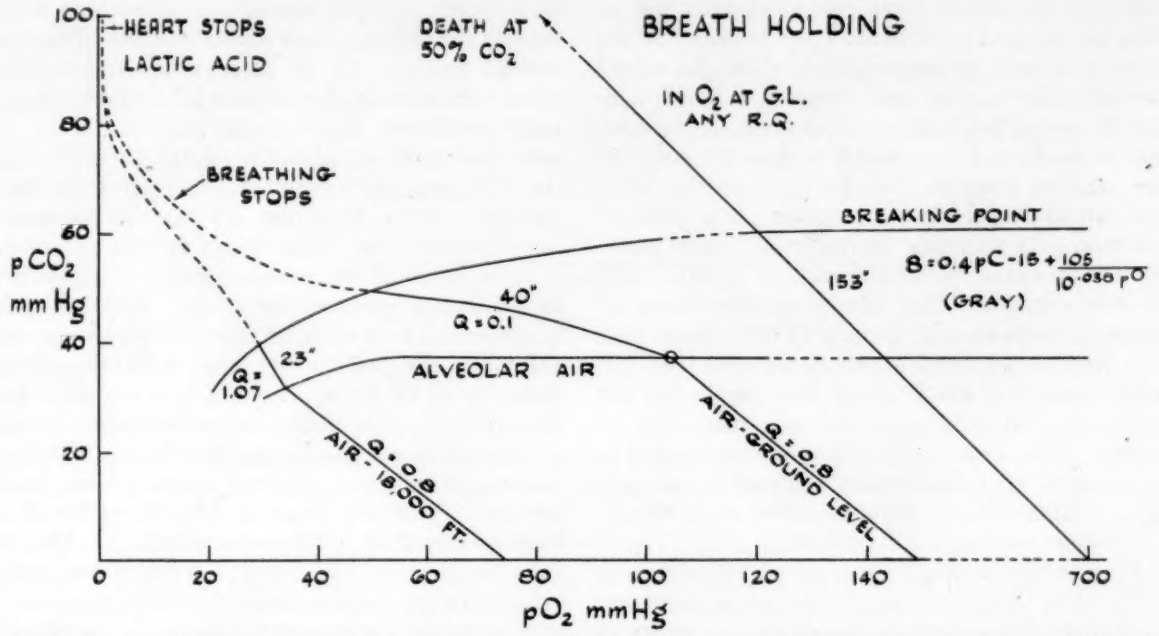

FIG. 4. Breath holding in pure O_2 at ground level (right), in air at ground level (middle) and at 18,000 feet simulated altitude breathing air (left). Ordinates and abscissae represent the pressures of CO_2 and O_2 respectively in the alveolar air in the lungs. Breath holding begins in each case at the alveolar air line and terminates at the breaking point line.

to the alkaline reserve of the body which combines with the carbonic acid when the concentration of CO_2 rises in the lungs. Experiments have shown that 1 liter of the body holds about as much carbon dioxide for 1 mm increase of pressure as 2.5 liters of gas volume in the lungs.[3] A 70 Kg man therefore can dispose of CO_2 in his body as well as if he were rebreathing 175 liters of air. Actually only about one-sixth of this equivalent volume is available during a breath holding experiment (with air at ground level) because the circulation does not have time to carry the CO_2 to all parts of the body.

The third breath holding pathway indicated is that at an altitude of 18,000 feet where the breath holding time is 23 seconds or less and most subjects come near fainting when the breaking point is reached. In this drastic experiment the R.Q. is high instead of low and the pathway on this chart is very steep. This is because the oxygen tension in the lung is so low that it promptly approaches the level of the venous blood so that no more O_2 can be taken up although the CO_2 is given off at the same rate as before. The CO_2 concentration at the breaking point is less than before because the stimulus due to low oxygen is greater and the sum of the two must add up, according to Gray's equation, to a constant value of about 8.

After the breaking point is reached the dotted lines show the probable further course which the alveolar air would follow if intake of air were prevented as in drowning or strangulation. These curves are based upon data (unpublished) obtained in my laboratory by Miss Florence Herber in experiments on anesthetized dogs. A few minutes after obstruction of the airway the breathing stops. Thereafter there is a continued rise in CO_2 tension as the oxygen falls to zero and the heart stops. Most of this terminal CO_2 must come from bicarbonate from which it is liberated by the lactic acid which forms in the muscles in considerable amounts and leaks into the blood. The CO_2 actually present in the muscles at death is 3 ml. per 100 gms. below normal in spite of the high CO_2 tension.

The terminal events when breathing pure oxygen at ground level as shown at the right in figure 4 are of special interest and of some practical importance. Consciousness is lost soon after the breaking point line is passed (if inhalation is prevented) and breathing movements may stop. In this case

[3] This figure is not exact and will be considerably less than this for high pressures of CO_2.

however the animal does not necessarily die so long as the glottis remains open because as the oxygen is used in the lungs more oxygen enters through the trachea and keeps the blood adequately oxygenated. CO_2 does continue to form but so much of it is retained within the body by the alkaline reserves that the CO_2 tension does not exceed 50 per cent of an atmosphere after 45 minutes. At this time the body of a man would contain an extra 14 liters of carbon dioxide, most of it in solution. This interesting experiment of Roth, Whitehead, and Draper (1947) shows that CO_2 itself is not lethal at the 10 per cent level but when breathing stops in air the oxygen in the lungs is soon exhausted and the man dies of anoxia. The exact point where apnoea begins is not known. A concentration of about 50 per cent CO_2 is lethal after a suitable period even though the oxygen supply is still sufficient.

Previous illustrations have shown two different kinds of R.Q. diagonals. In the experiments of Paul Bert the diagonals represent the progressive changes in the composition of the air of the closed vessel in which the animal was breathing. With an R.Q. of 0.6 the CO_2 goes up 6 mm when the O_2 goes down 10 mm (broken lines, figure 5).

The other kinds of R.Q. diagonal represents the composition of the *alveolar air* as a function of different rates of ventilation. Breathing more rapidly the alveolar point moves down the diagonal toward the inspired air point which corresponds to an infinite rate of ventilation. In these diagonals (solid lines, fig. 5), for an R.Q. of 0.6, the O_2 goes down only about 9 mm (instead of 10) when the CO_2 goes up 6 mm. The correct R.Q. lines can be plotted from the alveolar air equation shown in the slide. The reason for this difference is to be found in the volume change which occurs in the limited space of the lungs. When oxygen is consumed in excess of the CO_2 given out the volume of the gases in the lungs diminish and air comes in to fill the space. Thus, if the chest expands 1 liter, more than 1 liter of air comes in and in general more gas is inhaled than is exhaled. This extra inflow of air brings with it some fresh oxygen so that the resulting oxygen tension is a little greater than it otherwise would be. One of the advantages of this diagram is that it automatically takes care of complicated volume changes of this sort and the resulting dilutions. In Bert's experiments with closed vessels no fresh air could enter and the overall pressure must have fallen steadily. When the R.Q. is 1.0 both sets of diagonals are identical because there is no volume change. Likewise at any constant pressure or altitude where pure oxygen is inhaled the composition of the gases in the alveolar air must change

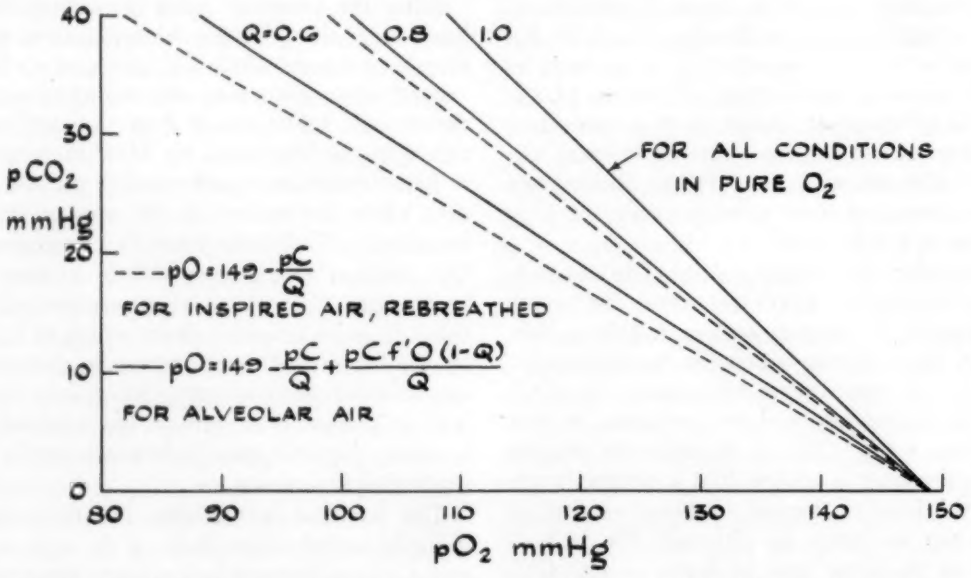

FIG. 5. The R.Q. diagonals for the inspired air in a closed vessel (broken) and for the alveolar air (solid) as indicated by the equations. The difference between two diagonals of the same CO_2/O_2 ratio (Q) is due to the volume change in the lungs and its magnitude depends upon the fraction of nitrogen in the inspired air.

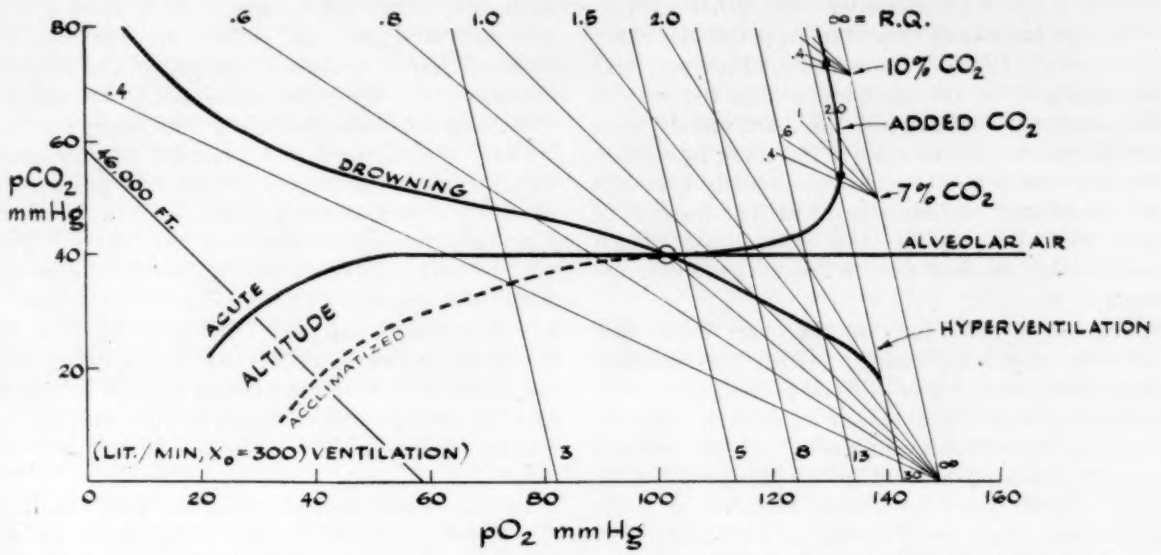

FIG. 6. The O_2-CO_2 diagram for ground level breathing air on the right and for 46,000 feet breathing pure O_2 in the left showing how all the R.Q. diagonals and isoventilation lines become reduced, in pure O_2, to a single 45 degree diagonal. The heavy lines represent movements of the alveolar point (represented by a circle) under different conditions.

along the diagonal marked R.Q. = 1 because, under such conditions, the sum of the pressures of these two gases must be equal to the dry gas pressure (B-47); a decrease in CO_2 of 10 mm must always be accompanied by an increase in O_2 of 10 mm or vice versa. Note that the alveolar air equation given in figure 5 reduces to $pO = 149 - pC$ when f'O (the fraction of O_2 in the inspired air) = 1.0. This is the equation for the R.Q. = 1.0 line.

The next chart (fig. 6) illustrates further some important features of the oxygen-dioxide diagram. Ordinates and abscissae represent as before the concentrations or pressures of these two gases in the alveolar air. At 40 mm of CO_2 and 100 mm of O_2 is a point marked by a circle which represents the normal alveolar air at ground level breathing air. This is located on the alveolar air line which shows how the CO_2 concentration of the alveolar air changes as the oxygen content of the inspired air diminishes, as it does with altitude, for example. As illustrated in previous charts the alveolar point lies on a diagonal which represents the normal ratio (R.Q. = 0.8) of CO_2 formed in metabolism to the oxygen used up. For every R.Q. there is a diagonal of a different slope all radiating from the point which represents the composition of the inspired or the ambient air. This whole family of radiating R.Q. lines must be thought of as shifting about on the diagram as the point of origin or the inspired air point shifts. If therefore one goes to high altitudes and the pO_2 diminishes the R.Q. diagonals are shifted along the axis of abscissae to the left. If the CO_2 concentration of the inspired air increases then the diagonals are moved upwards correspondingly. Such a set of diagonals forms a background grid against which the movements of the alveolar point can be read and interpreted.

Another part of this same grid is formed by the isoventilation lines. Such lines are drawn for 3, 5, 8, 13, 30, and infinite liters of air breathed per minute. All these lines are parallel to one another and all are to be shifted to the left along with the R.Q. diagonals in proportion to a decrease in the oxygen content of the inspired air. If the rate of ventilation and the R.Q. are known then the alveolar point must be represented by the intersection of the appropriate R.Q. diagonal and isoventilation line and the corresponding pCO_2 and pO_2 values can then be read off. On this chart it can be seen that the ventilation rate is a little less than 5 liters per minute. All the isoventilation lines are calculated for a given rate of oxygen consumption (300 cc per min.). They all have a slight negative slope which is equal to the fraction of oxygen in the inspired air. The mathematical expression which defines the isoventilation

lines and the R.Q. diagonals are rather complicated and have been described in detail elsewhere (Fenn *et al.*, 1946). Isoventilation lines may also be calculated on the assumption that the rate of CO_2 output is constant and these are merely horizontal lines across the chart. All these isoventilation lines refer to the air which actually gets into the alveoli and excludes therefore that fraction of each breath which fills the dead space (nasopharynx and trachea, etc.) where no gas exchange occurs.

The heavy black lines in the chart show four "alveolar pathways" along which the alveolar point can move under different conditions. In the lower right hand corner is such a pathway resulting from hyperventilation. If the subject merely increases his breathing rate voluntarily from 5 to 30 liters per minute his alveolar point must move from the 5 to the 30 liter isoventilation line. If the R.Q. remains at 0.8 the alveolar point would be at the intersection of this isoventilation line with the R.Q. = 0.8 diagonal. At that point however the pCO_2 is so low that CO_2 escapes from the stores in the body in excessive amounts so that the R.Q. increases temporarily to 2.0 as indicated in the diagram. In the course of half an hour, if the ventilation rate is maintained, it then gradually falls to the 0.8 line as the excess CO_2 is blown off.

It is interesting to find that this hyperventilation pathway seems to be continuous with the breath holding or drowning pathway already described. In the former the concentration of CO_2 in the alveoli is prevented from falling too far by release of CO_2 from the CO_2 stores of the body and in the latter, extra CO_2 is stored by the body to prevent the alveolar CO_2 concentration from rising too high. It should also be noted that in the process of hyperventilation the alveolar point moves to higher oxygen tensions and this increase in oxygen tension is further favored by the increase in the R.Q. This explains the improvement which the aviator experiences when he instinctively hyperventilates at high altitudes in response to the lowered concentration of oxygen in his lungs. The improvement is greater and more noticeable when air is breathed than with pure oxygen because of this "fanning out" of the R.Q. lines.

In the upper right corner of the diagram is shown the curve which represents the alveolar points which would result if the percentage of carbon dioxide in the inspired air were gradually increased, as by the escape of CO_2 from a high pressure tank into the room. In this case the point representing the gas tensions of the inspired air rises along the curve marked R.Q. = infinity. The oxygen tension gradually falls along this line because the oxygen is diluted by the incoming CO_2. For each successive inspired air point along the R.Q. = ∞ line there could be drawn a set of diagonals similar to those of the chart. Two sets of such diagonals are drawn at the points where the inspired CO_2 was 7 and 10 per cent of an atmosphere and the corresponding alveolar points are indicated on the 0.6 and 0.4 diagonals respectively.[4] At 10 per cent CO_2 the breathing rate has increased to 50 liters per minute and the alveolar point is very close to the inspired air point.

In this experiment it is important to note that the alveolar oxygen tension is increased as the CO_2 increases because of the greater rate of ventilation. For this reason the intake of CO_2 at high altitudes is of some benefit (although a similar intake of extra O_2 is better). If however CO_2 is added to pure O_2 for inhalation at altitude, the CO_2 tension and the ventilation increase exactly as indicated in the chart for air but the oxygen tension decreases *pari passu* with the increase in CO_2 because in this case the sum of pO_2 and pCO_2 must always be equal. When CO_2 increases, O_2 must therefore decrease an equal amount. Attempts to demonstrate a benefit from adding CO_2 at altitude under such conditions (absence of nitrogen in the gas inspired) must eventually prove abortive.

The chart which I have described with its background grid of R.Q. diagonals and isoventilation lines is particularly useful when dealing with problems involving the inhalation of air or nitrogen-containing mixtures for the presence of nitrogen introduces interesting complications. The situation is much simpler when pure oxygen is breathed as already indicated in figure 5. As the fraction of oxygen in the inspired air increases the R.Q. diagonals progressively close up like a fan until in pure oxygen they coincide with the R.Q. = 1.0 diagonal. Likewise the isoventilation lines gradually pivot around their points of intersection with the R.Q. = 1.0 line until they also coincide with it. This applies only to the isoventilation lines calculated on the assumption that the rate of oxygen consumption re-

[4] These points and other points on this curve were calculated from an equation of Gray (1946).

mains constant. This is rational because when pure oxygen is inhaled the rate of oxygen consumption can have no effect upon the composition of the alveolar air; for evey molecule of oxygen taken up in the lungs another molecule of oxygen will enter through the trachea. When the chest expands 1 liter more than 1 liter of oxygen will enter the lungs if the R.Q. is less than 1.0.[5] The isoventilation lines calculated for a constant rate of CO_2 output are not affected by changes in the oxygen content of the inspired air. When breathing pure oxygen, therefore, it is only the rate of ventilation and the rate of CO_2 output which determine the composition of the alveolar air.

These considerations show that the status of an aviator at high altitudes breathing pure oxygen can be represented as shown on the left of the chart in figure 6 by a single diagonal. The alveolar point must lie somewhere on this diagonal. For the particular altitude of 46,000 feet selected, the diagonal intercepts the axes at 59 mm this being the total pressure available (B-47 or barometric minus water vapor) for partition between CO_2 and O_2. If the ventilation rate remained at 5 liters per minute the alveolar pCO_2 would have been 40, as at sea level, and the alveolar pO_2 only 19 mm. If the ventilation rate were doubled, the alveolar pCO_2 would have been half as great or 20 mm and the alveolar O_2 would then be 39 mm. The problem of the respiration then is to breathe at such a rate that this pressure (B-47) is divided most advantageously between oxygen and carbon dioxide or so that the composition of the alveolar is as near as possible to the "normal range" indicated in figure 1. Our experiments have lead us to believe that on the average at high altitudes we instinctively breathe nearly but perhaps not quite fast enough for optimal performance.

This applies to an acute exposure to high altitudes like that obtained in an airplane. If one remains indefinitely at that altitude the rate of breathing continues to increase for a week or more as the sensitivity to carbon dioxide increases. This is shown by the dotted alveolar line for acclimatized men which was kindly supplied to me by my colleague, Dr. Rahn. This was drawn partly from his own data (obtained with Dr. Otis in Wyoming) and partly from the data of other mountain expeditions.

[5] This somewhat paradoxical statement is true if one considers the decrease in volume which occurs in the lungs before this liter is exhaled.

Another feature of acclimatization is the increased number of red blood corpuscles in the blood. In our laboratory Mr. R. T. Clark has recently shown that mice can tolerate exposures to high altitudes much better if they have been kept previously for a week or more in an atmosphere containing small amounts of carbon monoxide for this also increases the number of corpuscles. Another fact of extraordinary interest which seems to be appearing in his experiments is the fact that the CO in the closed vessel in which the mice are kept tends to disappear at an easily measurable rate which cannot be accounted for by combination with hemoglobin. This does not happen after the mice are removed. This experiment seems to confirm an observation which I made years ago in experiments with excised tissues which showed that CO can be burned to CO_2 by the oxidative enzymes existing in muscle.

The few examples which I have time to put before you could be multiplied almost indefinitely. In our laboratory for example we have devised an automatic device which continuously draws off the last of each breath and analyzes it electrically for both oxygen and carbon dioxide. In this way it is not difficult to obtain continuous records of the alveolar pathways in normal and pathological subjects when engaging in a great variety of respiratory experiments. These records are proving of great interest to the clinicians as well as to the physiologist. Eventually we hope to develop this into a simple instrument for bedside diagnosis. The complete interpretation of such experiments would require a rather complicated mathematical analysis but this is very much facilitated by the use of the graphical methods which I have tried to describe for you.

Man has been breathing just as he does today since the dawn of science. During that long time he has developed innumerable intricate machines the behavior of which has been analyzed mathematically in complete detail. A small fraction of this attention devoted to the understanding of the respiratory machinery would have elucidated that mechanism with equal perfection. It was not, however, until the exigencies of the recent war turned the attention of many minds to respiratory problems of high altitude flying that anyone clearly perceived the need for a strict mathematical treatment of the dynamics of pulmonary ventilation. Now through the independent efforts of several different laboratories we have at least an alveolar air equation and a ventilation equation

which define the steady state and these are nicely represented on our chart by convenient straight lines. Much more remains to be done in defining unsteady states and correlating respiratory and circulatory dynamics but it is hoped that what has already been accomplished represents in a small way a somewhat belated "promotion of useful knowledge."

REFERENCES

BERT, PAUL. 1943. Barometric pressure. Tr. by M. A. Hitchock and F. A. Hitchcock. Columbus, Ohio, College Book Co.

BEHNKE, A. R., R. M. THOMSON, and E. PREBLE MOTLEY. 1935. The psychological effects from breathing air at 4 atmospheres pressure. *Amer. Jour. Physiol.* **112**: 554–558.

CASE, E. W., and J. B. S. HALDANE. 1941. Human physiology under high pressure. I. Effects of nitrogen, carbon dioxide and cold. *Jour. Hyg.* **41**: 225–249.

CASS, RUTH E. 1947. Effects of high oxygen tensions upon the carbon dioxide production of skeletal muscle and other tissues of the frog. *Amer. Jour. Physiol.* **148**: 490–506.

ELLSBERG, E. 1939. Men under the sea. N. Y., Dodd Mead.

FENN, WALLACE O., HERMANN RAHN, and ARTHUR B. OTIS. 1946. A theoretical study of the composition of the alveolar air at altitude. *Amer. Jour. Physiol.* **146**: 637–653.

GRAY, J. S. 1946. The multiple factor theory of the control of respiratory ventilation. *Science* **103**: 739–744.

MACKAY, J. L. 1947. Effects of a narcotic level of carbon dioxide on the plasma potassium and respiration of cats. *Amer. Jour. Physiol.* **151**: 469–478.

RAHN, H., A. B. OTIS, and W. O. FENN. 1947. Alveolar gas changes during breath holding. *Fed. Proc. Exp. Biol.* **6**: 183.

ROTH, L. W., R. W. WHITEHEAD, and W. B. DRAPER. 1947. Studies on diffusion respiration. II. Survival of the dog following a prolonged period of respiratory arrest. *Anesthesiology* **8**: 294–302.

SEEVERS, M. H., 1944. The narcotic properties of carbon dioxide. *N. Y. State Jour. of Med.* **44**: 597–602.

STADIE, W. C., B. C. RIGGS, and N. HANGAARD. 1945. Oxygen poisoning. III. The effect of high oxygen pressures upon metabolism of brain. *Jour. Biol. Chem.* **160**: 191–208.

MARINE MOLLUSCAN PROVINCES OF WESTERN NORTH AMERICA: A CRITIQUE AND A NEW ANALYSIS[1]

IRWIN M. NEWELL

Department of Biology, University of Oregon

ABSTRACT

The "midpoints method" advanced by Schenck and Keen for defining the ranges and limits of marine zoogeographic provinces is not valid, since it is based on the unproved and false assumption that the majority of short range species which characterize a province are concentrated at the center of the province. The overlap area between the Oregonian and Californian Provinces proposed by Schenck and Keen is actually the most stable part of the coast faunistically, while the center of their "Californian Province" is the region of most extreme change, which is diametrically opposed to logical expectations.

Rate of change of fauna at successive points along a coast is the most reliable primary criterion for detecting faunal breaks and therefore the provinces which lie between them. Using Keen's data, the limits of the Oregonian Province are shown graphically to lie near 34° and 55° north latitude, the nodal points being at 34° 40′ and 55° 40′. To the south and north, respectively, of these points, the fauna is dominated by species of the Californian and Bering Provinces, while between the nodal points the Oregonian species outnumber the others.

The end points of the long range species are most numerous at the nodal points, and the short range species are likewise concentrated there. The midpoints of the long and short range species are out of phase by half a province, the midpoints of the former being most abundant in the center of the province, while those of the latter form maxima near the nodal points. It is believed that the concentration of single point records for Mollusca at the nodal points is due to a large extent to variability resulting from environmental conditions and hybridization.

INTRODUCTION

THE fixing of the boundaries of biotic provinces provides some of the most interesting problems of biogeography, but unfortunately has led to at least as much difference of opinion as agreement. The divergent views concerning the limits of biotic areas have their origin in two major deficiencies: inadequate data and a lack of methods for analysis of the available data. The need for analytical methods is great, for in the past our concepts of faunal limits have been developed largely from qualitative considerations of animal distribution, and consequently show errors resulting from the subjective limitations of the observer and his generally limited facts. Rough qualitative procedures have already yielded about as much to zoogeography as they are capable of, and future efforts should be turned in the direction of more precise quantitative attacks on the problem of animal distribution. This is not easy, for distribution is four dimensional, or at least three demensional if we restrict our efforts to extant species and faunae, and ignore their historical development. The difficulty of integrating all three dimensions for the components of a given fauna has undoubtedly deterred many from studies of this type. In marine zoogeography this difficulty can be partially resolved by considering the littoral forms distributed along the continental shelf of a simple coastline such as the Pacific coast of the Americas, for here there are only two major dimensions involved—length and depth. Or better yet studies can be confined to the intertidal zone, as Stephenson (1939, 1944) has done for the fauna of South Africa, where essentially only one dimension is involved, namely length. The development of methods of analysis can be simplified by beginning with these relatively less complex faunae.

H. G. Schenck and A. M. Keen (1936) advanced a method of analysis based on midpoints of ranges which is claimed to be valid for the delimitation of marine faunal provinces and the overlap areas between them. Little attempt will be made here to summarize the work of the writers noted above, since the analytical methods expounded by them do not lend themselves well to summarizing. Briefly, the "midpoints method" consists of plotting the number of midpoints of ranges of molluscs which fall at each latitude. No selection of any kind was exercised; even the forms known from single localities were included, these single point records being treated as midpoints. When this was done (Schenck and Keen, 1937: 925), two major peaks were found, one at 32–36°, another at 54–58° north latitude, on the Pacific Coast of North America. Following a line of reasoning which will be treated below, it was

[1] The writer was aided in this study by a faculty research grant from the Graduate School of the University of Oregon.

concluded that these midpoint maxima represent the centers of faunal provinces and that the provinces are separated by overlap areas in which fewer midpoints are found. The reader is referred to Schenck and Keen (1936, 1937, and 1940) and Keen (1937) for a further account of the development of the technique. My remarks here will be directed toward a critical review of some of the more unusual and questionable points raised by these papers, and also toward a presentation of new graphical methods of analysis of distribution data in the study of marine littoral zoogeography.

SCHENCK AND KEEN'S CONCEPTS OF FAUNAL PROVINCES

The concepts of Schenck and Keen regarding the nature of littoral zoogeographic provinces are exemplified by the following excerpt (numbering mine):

> (1) A marine molluscan province is a subdivision of a region, subregion or realm, populated by a distinctive assemblage of species. . . . (2) A province has at its center a high percentage of species with restricted ranges; the periphery, on the contrary, is marked by species of wide distribution. (3) The size of a province being variable, its limits are best defined by statistical methods. (Schenck and Keen, 1936: 923.)

The above definition (1) of a province is in general satisfactory, although it does not explain much and suffers from the introduction of three undefined terms (region, subregion, realm). It seems fruitless to attempt to define the term province, since one province can be "defined" only in terms of other provinces—an obviously illogical situation so far as definitions are concerned. Therefore, we shall probably have to content ourselves with a *concept* of the nature of a province, which despite its lack of finality would be preferable to an undefining "definition." In statement (2) we find the cardinal idea upon which the "midpoints method" of analysis of provinces is based and it is evident that this is strictly hypothetical, and no attempt whatever has been made to prove it in consistence with elementary logic. Not only is it unproved, but it is virtually certain that it could never be demonstrated in any province in nature. It might hold true in a perfectly idealized earth in which current systems, isotherms, and all other ecological factors were perfectly symmetrical in their distribution, and in which evolution was perfectly parallel on the two sides of the equator. Even on such an idealized earth, statement (2) above could hold true for only the tropical belt. If we assume distinct temperate provinces of equal magnitude on the two sides of the tropical province, the faunal center for these would not coincide with the geographical center. In fact, if the present rules regarding the relatively greater number of species in warmer waters held true, then there *could be no faunal center* as a study of figure 1 will demonstrate. In that figure two adjacent hypothetical faunal provinces are shown, each characterized by species, the ranges of which are indicated by the horizontal lines. Eurythermal species which might occupy both provinces have been omitted. The *tropical* province at the right is Schenck and Keen's (1936: 928) "hypothetical faunal province"; that at the left is a hypothetical *temperate* province constructed in the same manner but with a correction introduced in accordance with presently known principles of animal distribution, that is with the maximum number of species ranging equatorward from the geographical center of the province. The complete non-existence of a faunal center of the province as defined by midpoints of ranges is readily apparent. Thus the second point, and the cardinal point upon which Schenck and Keen's concept of a faunal province is based, is shown to be false even so far as an impossibly idealized temperate fauna is concerned. When we consider the many asymmetries in the ecological gradients which determine the shapes and sizes of faunal provinces in nature, then we cannot accept statement (2) above as even a rough approximation of actual conditions.

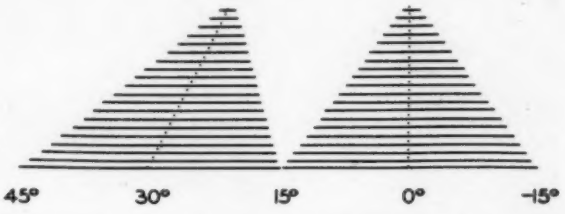

FIG. 1. Hypothetical tropical (right) and north temperate (left) faunal provinces constructed in accordance with Schenck and Keen's concept, but with the temperate fauna corrected to illustrate the general principle of the greater number of species south of the geographic center. In the temperate province, note the absence of a provincial center as defined by midpoints.

With regard to the desirability of delimiting faunal provinces statistically [statement (3)

above] the writer wishes to register complete agreement, but only with the reservation that the statistical methods employed must be based upon demonstrable ecological principles. The "midpoints method" is not, and must therefore be abandoned as a tool in zoogeographic studies of this nature. One of the greatest defects of this method is that *the midpoint of the range of a species is utterly devoid of any intrinsic biological significance*. It has significance of only an extrinsic type, derived from the fact that, by definition, it lies halfway between the end points which alone are of primary biological import. The range of a species is determined by ecological factors which become limiting at the margins of the range, and not at the midpoint.

PACIFIC COAST FAUNAL BOUNDARIES

Unless it could be shown that Schenck and Keen's concepts of a province and their methods of analysis lead to erroneous results, it would not be possible to reject their conclusions, regardless of how unsound they appear to be theoretically. Fortunately this can be done, thanks to the commendable and painstaking compilation of distribution records of western North American molluscs by A. M. Keen (1937). First, however, conclusions of previous workers are worthy of summary. Schenck and Keen have done this in the concise form of a map of the Pacific coast of North America beside which are graphed the faunal areas proposed by Dana, Woodward, Fischer, Dall, J. P. Smith, Bartsch, and Schenck and Keen. This has been published twice (1936: 932; 1940: 380) and will not be repeated here, especially since it would be difficult to improve upon the presentation by Schenck and Keen. The reader is also referred to the above papers for the citations of the references. It must be remembered that only the latter two authors have attempted to employ any systematic method of analysis in determining the ranges of the various areas; the others have simply presented views which were for the most part based solely on wide acquaintance with Pacific Mollusca. Nevertheless, there are significant points of general agreement. All but Smith and Dall recognized a major boundary at 23° N. latitude (Cape San Lucas, Lower California), while only Dana failed to note a break at 60° N. latitude. With the exception of Woodward and also Schenck and Keen, all are in agreement on the presence of a boundary of some magnitude at 32° 45′ to 34° 30′ N. latitude (Point Conception, California). Woodward did not recognize any break between 23° and 48° N. latitude.

The Californian Province of Schenck and Keen (1937: 931) extends from 48° to 23°, in other words directly astride the important boundary at about 33–34° noted by all other workers but Woodward. Therefore, what Schenck and Keen consider to be the center of their Californian Province coincides with the boundary between two adjacent provinces according to Dana, Fischer, Dall, Smith, and Bartsch. These diametrically opposed views clearly require closer scrutiny, for the views of men of such experience as the latter are entitled to more than summary dismissal, despite their failure to use statistical methods in arriving at their conclusions. It is noteworthy that Schenck and Keen qualify their Californian Province as follows:

> Between Cape Flattery (48° N.) and Cape San Lucas (23° N.)—our Californian province—no distinct assemblages can be recognized *provided no attention is paid to bathymetric distribution:* If, however, shore collections or shallow dredgings alone were considered, a fairly sharp line of division would appear in the vicinity of Point Conception.

The writer feels that the break near Point Conception is no mere statistical accident but is one of the most important faunal breaks along the entire coast, and demands more than a qualifying statement. Since there is obvious disagreement between the views of Schenck and Keen and those of previous workers regarding the Californian province, it is necessary to reexamine the available data to see whether the region of Point Conception is the center of a faunal province or is the boundary between two adjacent provinces.

THE NATURE OF FAUNAL PROVINCES AND OVERLAP AREAS

Having seen that the concept of a temperature province as a distribution area in which short-ranging species are found predominantly toward the center has no foundation even in theory, we must develop another concept. The one which is advanced here is by no means new, since it clearly has entered into the thinking of a number of zoogeographers. Within a zoogeographic province, significant numbers of the component species are restricted to the area. Most of these species are probably capable of surviving in other provinces with similar ecological characteristics, but are not found there because they are not favored

by ready means of dispersal, or are unable to withstand the conditions they would encounter during dispersal from their native province to other provinces. Since two adjacent provinces, if they are to be recognized as such on the basis of the above concept, must each contain significant numbers of characteristic species, then it follows that the two must be separated by a boundary area or "break" in which some species of the two provinces are found together. These are sometimes called overlap areas or transition zones, and are almost never sharp, since the transition from one zone to another is gradual and may cover several degrees. The transition consists of a dropping out of the species of one province, and a progressive increase of the species of the other province as one moves through the overlap area. Not all the species are affected by the transition zone, of course, for many of them are able to withstand a wider range of conditions and therefore occupy a wider geographical range. Nevertheless, any region in which there is a marked increase in the rate of addition and loss of species as one proceeds through the region must be suspected of being an overlap area, since it is most unlikely that such an increase in rate of change would affect only the long ranging species and not the shorter ranging species which characterize the individual provinces.

Between the overlap areas we should expect to find the provinces themselves, in which the fauna is relatively stable and the rate of addition or loss of species is appreciably lower than in the overlap areas. Areas of pronounced change in fauna are almost invariably associated with major changes in ecological conditions such as temperature, currents, sharp changes in substrate such as the change from a rocky embayed coast to an open sandy coast, etc. Where there are no obvious ecological changes, more subtle ones should be suspected and looked for, or the possibility of statistical artifacts arising from concentrated collecting should be examined.

RATE OF CHANGE OF FAUNA

In figure 2 the rate of change of the fauna along the Pacific coast of North America from the equator to 72° N. latitude is graphed. This graph represents the simple equation: the rate of change at each latitude equals the total number of species with *either* northern or southern limit at that latitude, divided by the total fauna at the latitude. Species which are listed as occurring at a single

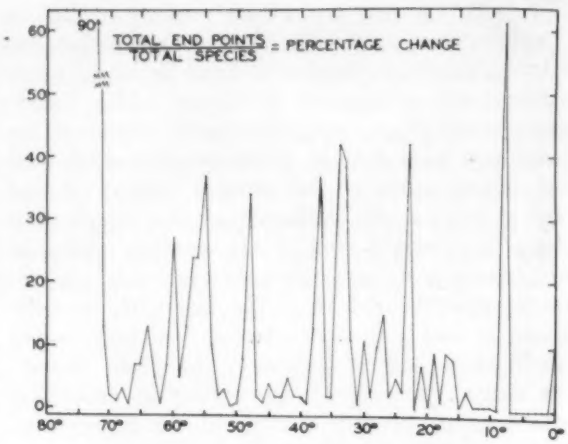

FIG. 2. Rate of change of fauna observed in scanning the coast in a north to south direction or vice versa. Points of maximum change occur at 32–34° N. and at 54–58° on the Pacific coast of North America. A 90 per cent change in fauna occurs at 72°, owing almost entirely to a dropping out of species.

point (e.g. 33°) in Keen's list (1937) are omitted. These would only serve to exaggerate the peaks, anyway, as one can readily see by examining figure 3, in which the number of single point records for each degree of latitude is shown. The peaks in figure 3 correspond precisely with those in figure 2, except at the northern limit, where statistical artifacts are probably operating owing to inadequate collecting, and south of 32°, where the records are not included in Keen's list. From figure 2 it is apparent that the points of maximum change in fauna occur at 8°, 23°, 33°, 34°, 37°, 48°, 55°, 56°, 60°, and 72° N. latitude. Of the breaks between Cape San Lucas and Puget Sound,

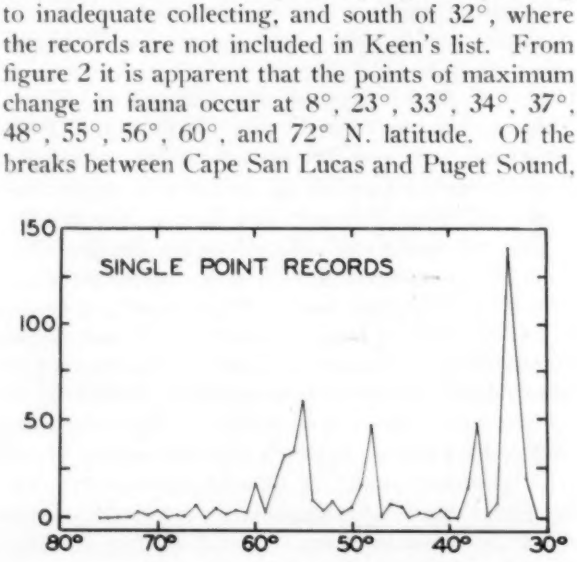

FIG. 3. Number of species known from only one locality according to Keen's list (1937). These are omitted from the present author's analysis, but the position of the maxima shows that this procedure has no effect on the conclusions.

by far the most significant one occurs at 33–34°, for here 39–42 per cent of the species found at these two latitudes are in the process of dropping out of the fauna or being added to it if we assume progression of the observer from south to north or north to south. In other words, 39 per cent of the species at 33° N. latitude have either their northern or southern limit there, while 42 per cent of those occurring at 34° find their northern or southern limit at that point. Therefore, the change in fauna at 33–34° is even more pronounced than one might think from first glance, since the total change here is divided between two different latitudes while the other peaks, with the exception of that at 55–56°, occupy only a single degree. Surely such a high rate of change as that at 33–34° and at 37° is scarcely to be expected in the *center* of a province, yet these points fall almost squarely in the center of Schenck and Keen's Californian "province." The present writer feels that this alone is ample evidence for rejecting the province established by the "midpoints method," and reinstating the boundary recognized by nearly all other workers—between 32° 45′ and 34° 30′.

If we turn our attention now to the "overlap" between the Californian and Aleutian provinces of Schenck and Keen we will find the equally remarkable fact that the interval between 39° and 47° *is the most stable interval of any magnitude along the entire coast.* For a distance of eight degrees the change at each degree ranges from 1 to 5 per cent, and averages only 2.3 per cent for this great distance! Schenck and Keen's "overlap area" extended from 42° to 48°. This, too, must be rejected since it is obvious that overlap areas must be areas of rapid change and not of extreme stability. For that part of the coast from which the most reliable data are available, the "midpoints method" for determining faunal provinces gives results which are incompatible with the results obtained by more direct means, and the method is evidently invalid for the present purpose.

The writer believes that *rate of change of fauna* provides the best index to the location of faunal provinces and breaks. There are, however, no exact rules that can be laid down; rather, each problem must be considered by itself, and the solution worked out empirically. Each point at which a large change occurs in the fauna must be suspected of being a boundary area and tested. However, a good idea of the importance of the several peaks of pronounced change in fauna which can be seen in figure 2 may be gained by comparing the relative areas subtended by the separate peaks, or by considering the average per cent change in fauna through the peak. When this is done, the results shown in table 1 are obtained.

TABLE 1

RELATIVE VALUES OF THE PEAKS OF GREATEST CHANGE IN MOLLUSC FAUNA OF WESTERN NORTH AMERICA
(Based on figure 2)

Peak	Center	Relative Area Subtended by Peak (no units)	Average Percentage Change in Fauna
7– 9°	8°	61	20%
22–24°	23°	44	15
31–35°	34°	94	19
36–39°	37°	51	13
47–50°	48°	41	11
53–59°	55°	112	17
59–62°	60°	37	10
70– °	72°	59	35

There is good reason to doubt the reliability of these figures for the peaks at 8° and 23°, for here the data are certainly incomplete. This is seen in figure 4 where the total mollusc fauna at all points south of 23° is less than that at 56°. It is certain that this does not represent true conditions, but is a reflection of the limitations of the data, for no species were included in Keen's check list which did not extend at least as far north as 32°.

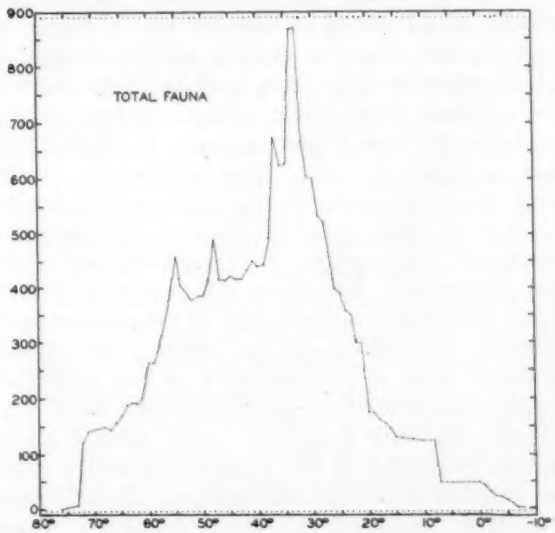

FIG. 4. Total species in the fauna at each latitude, exclusive of single point records. The drop in the fauna toward the south reflects the fact that Keen's list included only those species which reached at least as far north as 32°.

The peak centering at 72° cannot be completed since only seven species are recorded beyond 72°, so that the relative area given may be considered as representing roughly half the actual area, and the average percentage change is based on that part of the curve from 70° to 72°. It almost certainly represents a true zoogeographic boundary despite (and to a large extent because of) the paucity of records north of Point Barrow.

Two of the remaining peaks stand out above the rest, the first centering at 34°, the second at 55°, and these will be tested as possible provincial boundaries. For this purpose the ranges of all species of Pelecypoda, Gastropoda, and Scaphopoda were plotted on graph paper in the manner shown in figure 7 which represents the ranges of all Pelecypoda whose northern limits lie at 65 to 76°. Single point records are not included since they complicate the analysis unnecessarily and add nothing to the results.

PRELIMINARY ANALYSIS OF THE OREGONIAN PROVINCE

The first step in the analysis consists of counting the species in the fauna at the center of the area which is most likely to be a province, in the present case, 417 species whose range includes 44° N. latitude, halfway between 34° and 55°. Proceeding southward from this, the change in fauna is determined for each degree of latitude, in terms of number of the original 417 species which are still found at each successive degree of latitude, and also the number of "new" species which appear at each degree. When these figures are graphed, they produce two separate curves as shown by broken lines in figure 5, which intersect at about 34° 50′. This intersection is the most exact indication of the center of the break, for it is the point at which the fauna theoretically contains equal numbers of Oregonian and Californian species. Biologists will realize of course that a collection made precisely at this intersection, hereafter referred to as the *nodal point*, would probably show a predominance of one group of species over the other, depending upon local conditions. In other words, the nodal point is not a natural phenomenon but a statistical abstraction; nevertheless it is a reliable indicator of the faunal boundary. Surely it is preferable to the vague "overlap area" since it can be precisely determined and is not liable to extreme variability stemming from subjective judgment. Furthermore, the fixing of the nodal point aids in the detection of the factor or factors which are responsible for the rapid change in the composition of the fauna.

When the above process is repeated for the area north of 44°, the nodal point is found to lie at about 55° 30′. Here a problem arises in placing the nodal point, for there is no way of deciding from the available information whether the break occurs along the mainland or near the end of the Alaskan Peninsula, that is, whether the Gulf of Alaska (which extends to 61°) should be included in the Oregonian province. From a consideration of the currents in the Gulf of Alaska, it is most improbable that there is any significant break along the mainland coast. The faunal break here undoubtedly occurs near the end of the Alaskan Peninsula between 55° and 56°, and the Gulf of Alaska is a part of the Oregonian province.

It is significant that marked changes in the slope of the curves for both the 44° species and the "new" species occur at 37° and 48°. These inflections reflect important changes in the composition of the fauna at each of these two points, but the fauna here is still predominantly Oregonian in nature, there is no nodal point, no "break," and therefore, according to the writer's concept, no new province.

CORRECTION OF THE PRELIMINARY CURVES

A little reflection on the method by which the dotted or dashed curves in figure 5 were derived will reveal that these are only first approximations

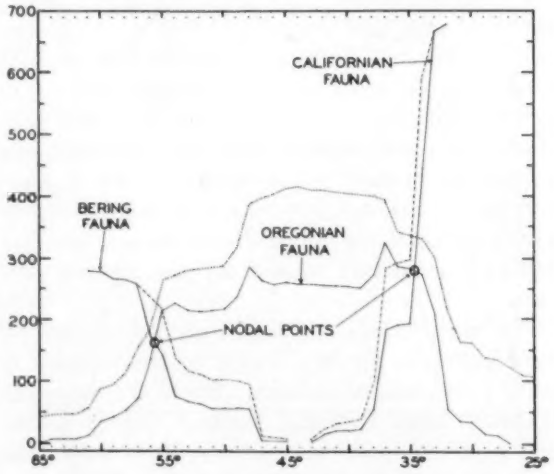

FIG. 5. Preliminary (dotted and dashed) and corrected (solid) curves for the Bering, Oregonian, and Californian faunae. The latter at least will require further correction when more is known of the distribution of the southerly species.

and must be corrected. In the first place the curve for the Oregonian species includes many long range eurythermal species, or perhaps deep water forms which cannot be attributed to the Oregonian province any more than to the Californian or Bering province. These should therefore be subtracted from the curve for the Oregonian species. In the second place not all the species which were included in the preliminary curves for the "new" species are really "new," for some belong to the Oregonian fauna. For example, a species which ranges from 42° to 37° would be counted as a "new" species in these preliminary curves, whereas it is strictly Oregonian in its distribution. Such species must be added to the curve for the Oregonian species. In the third place, these Oregonian species must be subtracted from the curves for the "new" species. The first and second corrections were made by selecting and replotting all records for the following three groups of species:

a. Those short range species whose total range lies between 43° and 34° inclusive and whose northern limit is not south of 36°.

b. Those short range species whose total range lies between 45° and 56°, and whose southern limit is not south of 54°.

c. Those long range species whose northern or southern limit does not exceed 63° or 28° respectively, and whose range includes 44°.

The first and second groups exclude any species whose range exceeds the nodal points by one degree or more, and it is also evident that the choice of the limits excludes those species with ranges of less than two degrees at the nodal points (34°–35°, 55°–56°). There is little question that all such short range forms are assignable to the Oregonian province. It is more difficult to know what tolerance should be allowed in the case of the long range species, but the exclusion of all forms which range as far as or beyond the middle of the neighboring province appears to be sufficiently exacting. All such species which range southward beyond the nodal point have 59 per cent or more of their total range within the Oregonian province, whereas the northerly ranging species have 58 per cent or more of their range within these limits. As thus defined, by far the majority of long range Oregonian species have significantly more than 60 per cent of their range within the Oregonian Province. In all, 84 per cent of the total distribution of the species here defined as Oregonian lies between the nodal points. An appreciable part if not all of the remaining 16 per cent is admissable overlapping. The writer feels that little if any of this 16 per cent can be criticized as error, and that all the species in the three groups above are logically assignable to the Oregonian province. When the total fauna at each degree of latitude is now plotted, the curve shown by the convex solid line in figure 5 is obtained.

The third correction noted above was executed by counting the species coming in at each latitude which were originally counted with the "new" species in drawing the preliminary curves (fig. 5), but which in reality are Oregonian species. These are the short range species which fall in groups *a* and *b* (see above). When these are subtracted from the preliminary (dashed) curves for the "new" species in figure 5, the sigmoid curves indicated by the solid lines are obtained. Although the corrected curve for "new" species is appreciably lower than the preliminary curve at the nodal points, there is only a shift of a few minutes of latitude in these points so that none of these corrections has any significant effect on the position of the nodal points.

On the Pacific Coast of North America we can therefore recognize three prominent faunal breaks embracing two well defined provinces. The breaks center at 72°, 56½° and 34½° N. latitude. The included provinces are the Bering Province and the Oregonian Province. The name Bering Province is introduced because there is no appropriate term that has yet been employed for the area delimited here. Dana recognized a Frigid Province with a southern limit at 56°, but the writer is not in favor of trying to name zoogeographic provinces in accordance with preconceived ideas of the meaning of such terms as frigid, temperate, arctic, etc. There are several faunal areas in the world which have equal claim to the name frigid, but which bear little resemblance to each other faunistically. Any zoogeographic classification of the earth can be based logically only on the faunae themselves and not on climatological or hydrographic features; hence regions with different faunae should be given different names, and this precludes the application of any general name to one province at the cost of excluding the others. The simplest way is to determine the limits of each province on the basis of as much of the fauna as possible, and then to name the province

after one of the more prominent geographic features associated with it.

The Oregonian Province delimited here coincides precisely only with the Oregonic of Bartsch, but the writer prefers to retain it in order to keep confusion to a minimum. Variations of the term Californian have been used by most authors, with variable opinions as to its limits, but the writer would retain the name Californian Province for that faunal area just south of the Oregonian.

If we must look for subdivisions of these, only one appears to stand out above the others, and that is what will be referred to here as the Columbian Subprovince (named for British Columbia). Its southern limit is at 48°, but its northern limit is not detectable with certainty on the basis of Keen's list because of the complexity introduced by the arched form of the coast around the Gulf of Alaska. Presumably, however, it extends to the nodal point at the end of the Alaskan Peninsula.

The southern limit of the Californian Province cannot be definitely established without a more complete knowledge of the fauna than we now have. Considering the fact that mangroves and coral reefs of at least modest dimensions are found south of Cape San Lucas (23°) and also in the Gulf of California, it is quite probable that Cape San Lucas is an important break. Whether the fauna between 23° and 34° meets the additional test of being characteristic cannot be stated at present, however.

FURTHER OBSERVATIONS ON THE CHARACTERISTICS OF FAUNAL PROVINCES

DISTRIBUTION OF THE LONG RANGE SPECIES

An illuminating demonstration of the reality of the faunal provinces and boundaries established above comes from the consideration of the distribution of the species whose ranges cover fifteen degrees, or more, of latitude. The results parallel those obtained with the total fauna in all respects, and prove that no errors are introduced by the inclusion of the short range species. The solid line in the upper part of figure 6 shows the total number of end points at each degree of latitude, the maximum numbers being found at 72°, 60°, 55°, 48°, and 37–34°. The peak at 60°, as was pointed out before, is due to the species which find their northern limit at that latitude, and probably has no great significance for the western coast of Alaska, but only for the northern coast of the

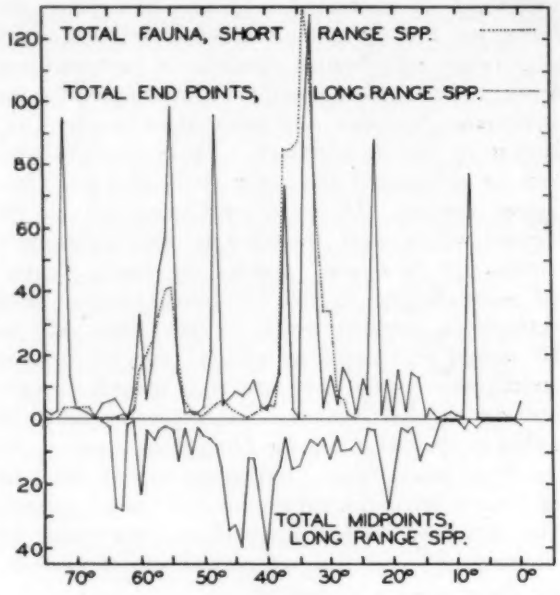

FIG. 6. The upper section shows the coincidence of the peaks of total fauna of short range species at each latitude with the great maxima of end points of long range species, and with the boundaries at 33–34° and 55–56°. The lower section shows the midpoint maxima for long range species falling near the centers of the provinces.

Gulf of Alaska. The break at 48° is undoubtedly real and must be associated with the sharp change in coast line from the straight, open coast found from California to the Strait of Juan de Fuca, to the deeply embayed coastline of British Columbia and southern Alaska, with its chain of offshore islands. The maximum change here affects the long range species. In figure 6 this is shown by the great discrepancy between the high peak representing the loss and gain in long range species and the small peak showing only five short range species at 48°. The significance of the coincidence of the peak at 33–34° with the nodal point requires no elaboration. There is a pronounced change at 37° also, but between 37° and 35° the fauna is still dominated by Oregonian species. As a group, the long range species, except for forms with extremely long ranges, probably give as good an indication of the position of faunal breaks as any other group.

DISTRIBUTION OF THE SHORT RANGE SPECIES

Here we are confronted with a situation which has no single explanation, and which has a number of rather important implications. The coinci-

dence of the peaks of abundance of the single point records (fig. 2) with the nodal points may be due to a very large extent to the listing of varieties and species without regard to their actual rank. For obvious reasons it was impossible for Keen to publish a check list which would be entirely accurate systematically, since this would require decades of revisional work, and there are undoubtedly many mere varieties which have been listed which really deserve no systematic ranking. The influence of conchology on molluscan systematics has its unfortunate aspects, and one of these has been the introduction of a large number of trivial varieties which should never have been named. In the genus *Odostomia* some 45 forms are listed by Keen as having been reported from single localities in the vicinity of 33–34°. The writer is not a malacologist and cannot say that any one of the 45 names which have been applied to these forms is invalid, but on the basis of general acquaintance with the group and a very specific acquaintance with the systematics of other groups, it is felt safe to say with no reservation that many of the single point records are for simple varieties which do not merit recognition.

The fact of their coincidence with the faunal breaks is a good indication of this. In the region of the breaks there are many species which are living under conditions which for them are limiting. It is to be expected that many individuals, while they may be able to survive, will develop abnormally. Shell characteristics, upon which the bulk of molluscan taxonomy at the species level rests, are particularly subject to environmental influence, and it is not unreasonable to expect that the greatest rate of aberration would be found at those points where conditions marginal for the greatest number of species exist. Hybridization probably accounts for a certain amount of the variability also, for it is precisely at the faunal boundaries that the maximum incidence of hybridization would be anticipated. Here the ranges of many allopatric species overlap slightly, providing abundant opportunity for hybridization.

The great variation found in the overlap areas is probably of this dual origin—ecological and genetic. Precisely what proportion of the forms listed from single localities are varieties or hybrids cannot be stated, of course. In all there are about 625 single point records, making up 32 per cent of the list. Of these, 46 per cent are found at 33°, 34° and 37°, and an additional 19 per cent at 55° and 56°. About 10 per cent occur at 48°, and the remaining 25 per cent are distributed among the other 39 latitudinal intervals between 32° and 76°.

The above observations also may be applicable to the 120–125 forms whose ranges cover no more than two degrees, although here there must be a higher percentage of good species. However, even if we consider only the species with ranges of four to five degrees, these too show a concentration around the nodal points, and it would be impossible to impugn these, since the majority are quite certainly good. A concentration of the short range species about the nodal points therefore appears to be a fact.

DISCORDANT DISTRIBUTION OF THE MIDPOINTS OF LONG AND SHORT RANGE SPECIES

If the provinces and nodal points delimited by the writer here are accepted, and the writer can see no reality in the province concept if they are not acceptable, then it is easy to demonstrate the incorrectness of Schenck and Keen's initial assumption that short ranging species typically are most abundant at the center of a province. In table 2 the number of midpoints of Oregonian species found at each degree is given, along with the length of range of the species whose ranges center at that latitude. Each species is represented by a number in that vertical column which represents the latitude at which the midpoint lies. The number itself gives the number of degrees of latitude occupied by the species. A study of this table shows these important facts: (1) the number of midpoints is greatest just north of the nodal point, (2) except for this peak at 35° the maximum number of midpoints is found toward the center of the province, (3) the short range species are most abundant at the periphery of the province. Even if we omit the 70 midpoints at 35° the distribution of midpoints is still skewed south of the faunal center, 145 falling at the ten points from 46° to 55°, and 202 from 36° to 45°. This is in agreement with the statement made previously that in an ideal temperate province we should expect to find the greatest number of species (and hence of midpoints) equatorward from the geographic center. Naturally the number of midpoints of long range species is greatest toward the center, but we cannot attach immediate significance to this fact since it may be due to the mode of selection of the Oregonian species, for if the range of a species was great, and its midpoint fell far to one side of the center of the

TABLE 2

DISTRIBUTION OF MIDPOINTS OF ALL OREGONIAN SPECIES
(Explanation in text)

55°	54°	53°	52°	51°	50°	49°	48°	47°	46°	45°	44°	43°	42°	41°	40°	39°	38°	37°	36°	35°
2	1	3	5	7	10	2	1	2	4	2	8	10	11	8	3	4	8	1	3	2
2	2	21	6	7	10	12	2	15	16	22	8	10	11	9	13	11	8	1	4	3
2	2	21	6	7	12	14	2	20	17	22	14	10	11	12	14	12	9	1	4	3
2	2	21	6	7	12	16	11	20	18	22	20	10	11	14	15	12	9	1	4	3
2	14		7	7	16	24	13	20	18	22	20	12	11	14	15	13	10	9	4	3
2	20		7	7	19		22	20	18	22	20	12	11	14	15	18	10	9	4	3
2			8	7	21		23	20	18	22	20	12	11	15	15	19	11	19	4	3
2			8	7	26		23	21	18	22	20	12	11	15	15	19	19		4	3
2			8	7			23	21	18	22	21	21	11	16	15	19	20		6	3
2			8	7				26	19	22	21	22	12	16	15	20	20		6	3
17			8	7				26	19	25	21	22	18	16	15		20		6	3
17			8	7				26	19	25	21	23	18	16	15				16	3
			8	7				26	20	25	21	23	19	16	15					3
			8	7				26	20	25	22	23	24	16	15					3
			8	7				26	22	26	22	23	24	16	15					3
			8	7					22	26	22	23	24	27	15					3
			8	8					22	30	22	23	25	27	15					3
			8	8					22	30	22	26	25	27	15					3
			18	20					23		22	27	25		15					3
			18						23		22	27	25		15					4
									23		22	27	25		15					4
									23		22	28	25		15					4
									24		22		28		15					4
									24		22		29		15					4
									24		22		29		15					4
									24		22				15					4
									24		22				15					4
									24		22				15					4
									24		22				15					4
									24		22				15					4
									24		22				16					4
									24		22				20					4
									24		22				20					4
									26		22									4
									26		23									+24
									26		23									4
									26		23									4
									27		23									5
									27		24									5
									27		25									5
									27		25									5
									27		25									5
									27		32									5
									28		32									5
									28											5
									28											5
									28											5

province, it would have been eliminated from consideration as an Oregonian species.

To eliminate this bias, it is necessary to consider the distribution of the midpoints of all long range species, and this has been done in the lower section of figure 6. Here it can be seen that for all species of Mollusca whose ranges occupy fifteen or more degrees of latitude, the midpoint maxima fall at 60–64° and 40–46°, and therefore near the center of the Bering and Oregonian Provinces. The dotted line in figure 6, while it represents total short range species at each latitude, can also serve to indicate the distribution of the midpoints of the short range species (the curve showing total

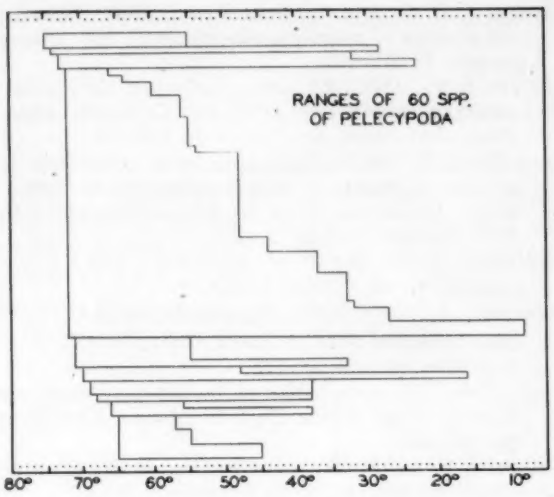

FIG. 7. Sample graph showing method of plotting distribution records for present analysis. The lines of the graph paper are not shown. The narrowest bars represent single species.

fauna is smoother, but the maxima and minima correspond precisely with the maxima and minima of the midpoints of the short range species). A comparison of the distribution of midpoints of long and short range species shows that these are out of phase by roughly half a province, the midpoint maxima for long range species falling in the centers of the provinces, and for short range species at the nodal points. This reveals another major defect of the "midpoints method" for Schenck and Keen failed to recognize this discordant distribution of the midpoints of the short and long range species.

CONCLUSION

The present study has been based on the same data used by Schenck and Keen in their statistical analysis of marine faunal provinces of western North America, but the conclusions are very different. The new analytical procedures presented here verify the views of several earlier workers, most notably Bartsch, who also recognized important faunal boundaries at 34½° and 56° N. latitude. It has been shown that the stability of the littoral fauna varies from region to region if one "scans" the coast in a north-south direction or vice versa, some regions showing a high rate of gain and loss of species to the fauna. Points of maximum change represent the overlap areas, while within the provinces stability is relatively high. Where the faunae of two adjacent provinces are equally represented, a nodal point is found. It is important to note that while the nodal point has a great deal of reality on graph paper its detection in nature may be more difficult, but it is no more difficult to detect than the vague "overlap areas," and the computation of its position aids in identifying the factors which are responsible for the faunal change.

The failure of the "midpoints method" stems principally from two errors, one of logic and one of observation. In the first place, Schenck and Keen attached a significance to midpoints which was entirely unproved (and false), and then proceeded to "disprove" the views of earlier students. In the second place they failed to observe that the distribution of midpoints is markedly different for long and short range species.

Finally it should be pointed out that Schenck and Keen have also employed midpoints in comparing collections of fossil Mollusca taken at various stations, principally with a view to determining what temperature conditions existed there at the time the organisms were alive (Schenck and Keen, 1939; Schenck, 1945). It should be clearly understood that the criticisms directed at the "midpoints method" for determining faunal provinces are in no way applicable to the use of midpoints for this type of paleoecological studies. For the latter purpose the use of midpoints is theoretically valid. However, some of the points brought forward in the present paper might be of value in the logical utilization of the method.

BIBLIOGRAPHY

BARTSCH, PAUL. 1912. A zoogeographical study based on the pyramidellid mollusks of the west coast of America. *Proc. U. S. Nation. Mus.* **42**: 297–349.

——. 1921. Ocean currents, the prime factor in distribution of marine mollusks on the west coast of the Americas. *Special Publ., Bernice P. Bishop Mus.* **7**: 505–526.

DALL, W. H. 1909. Report on a collection of shells from Peru, with a summary of the littoral marine Mollusca of the Peruvian zoological province. *Proc. U. S. Nation. Mus.* **37**: 147–294.

——. 1916. On some anomalies in geographic distribution of Pacific Coast Mollusca. *Proc. Nation. Acad. Sci.* **2**: 700–703.

——. 1921. Summary of the marine shellbearing mollusks of the northwest coast of America, from San Diego, California, to the Polar Sea. *Bull. U. S. Nation. Mus.* **112**: 1–217.

DANA, J. D. 1853. Crustacea: United States Exploring Expedition, part II, **14**: 1–1618.

EKMAN, SVEN. 1935. Tiergeographie des Meeres. Akademische Verlagsgesellschaft, Leipzig.

ESSENBERG, C. 1918. The factors controlling the distribution of the Polynoidae of the Pacific Coast of North America. *Univ. Calif. Publ. Zool.* **18** (11): 171–238.

FISCHER, P. 1887. Manuel de conchyliologie et de paléontologie conchyliogique ou histoire naturelle des mollusques vivants et fossiles. Paris.

HUTCHINS, LOUIS W. 1947. The bases for temperature zonation in geographical distribution. *Ecol. Monogr.* **17**: 325–335.

KEEN, A. M. 1937. An abridged check list and bibliography of west North American marine Mollusca. Stanford Univ. Press.

SCHENCK, H. G. 1945. Geological application of biometrical analysis of molluscan assemblages. *Jour. Paleont.* **19**: 504–521.

SCHENCK, H. G., and A. MYRA KEEN. 1936. Marine molluscan provinces of western North America. *Proc. Amer. Philos. Soc.* **76**: 921–938.

SCHENCK, H. G., and A. MYRA KEEN. 1937. An index method for comparing molluscan faunules. *Proc. Amer. Philos. Soc.* **77**: 161–182.

SCHENCK, H. G., and A. MYRA KEEN. 1940. Biometrical analysis of molluscan assemblages. *Soc. Biogeographie* **7**: 379–392.

SMITH, J. P. 1919. Climatic relations of the Tertiary and Quaternary faunas of the California region. *Proc. Calif. Acad. Sci.*, ser. 4, **9**: 123–173.

SOOT-RYEN, R. 1932. Pelecypoda, with a discussion of possible migrations of Arctic pelecypods in Tertiary times. Norwegian Polar Expedition, "*Maud*," 1918–1925, Scientific Results **5** (12).

STEINBECK, JOHN, and E. F. RICKETTS. 1941. Sea of Cortez. N. Y., Viking Press.

STEPHENSON, T. A. 1939. The constitution of the intertidal fauna and flora of South Africa, Part 1. *Jour. Linn. Soc.* **40**: 487–536.

——. 1944. The constitution of the intertidal fauna and flora of South Africa, Part 2. *Annals Natal Mus.* **10**: 261–358.

SVERDRUP, H. U., M. W. JOHNSON, and R. H. FLEMING. 1942. The oceans. N. Y., Prentice-Hall.

THE GROWTH OF PARTICLES IN SMOKES AND CLOUDS AND THE PRODUCTION OF SNOW FROM SUPERCOOLED CLOUDS*

IRVING LANGMUIR

Associate Director, Research Laboratory, General Electric Company

I THOUGHT, tonight, I would tell you about a number of different things concerning a series of researches that started during the early part of the war, in fact, about a year before we entered the war, and that carried through up to the present time. I want to leave with you certain threads of thought which I hope you will keep in mind as I tell you about these different things. There is a certain amount of continuity, although it may not seem so at first sight.

These threads are of many kinds. In the first place, the researches grew out of one another and yet there was no obvious connection. It was always an accidental connection.

Several years ago, Dr. Whitney introduced into the laboratory in Schenectady the word "Serendipity," and it is very appropriate for certain kinds of scientific research. The word was invented or coined by Sir Horace Walpole about two hundred years ago, in letters that he wrote to Sir Horace Mann. It started in a very trivial way. He was attending a dinner in which some discussion came up as to what the history was of a certain painting and nobody seemed to know. Finally he was able to discover it owing to fortuitous circumstances. He said that this was an example of what he liked to call serendipity. It was a word that he coined from a silly fairy tale that he had read as a child about the three princes of Serendib. Serendib was the old name for the island of Ceylon. These princes of Serendib had the happy faculty of making discoveries, apparently by accident, but always very fortunate discoveries, discoveries which they did not seek.

Now I would like to give my own definition of serendipity and see if you can find its application in the activities about which I shall tell you: It is the art of profiting from unexpected occurrences. It has a great deal to do with the planning of research.

The American public likes to think of supermen. There are supermen of all kinds: in Wall Street, in labor unions, in our Government. There are economic planners who can plan all things, and given time enough they would make everything very good or very bad. They are supposed to know how to plan. Now the examples I want to give you are all examples of unplanned research where things happened in a way that nobody could have planned and arranged and as a result we got results which were most satisfactory.

They occurred as though we were just drifting with the wind. These things came about by accident, and yet you will see that a series of rather fortunate accidents started us on a research problem just about the time we were finishing another, in a way which gave a continuity which has turned out to be very useful.

Another thread that goes through it all is this very idea of planning; as to whether planning is always successful in accomplishing the results; as to whether or not there are definite relations of cause and effect. If there are definite relations of cause and effect and you can see the effect produced by a given cause, then of course we should go ahead and plan.

But suppose in some cases there is no relation of cause and effect. Then planning does not get us very far. All we can do is, like serendipity, put ourselves into a favorable position to profit by unexpected occurrences.

Our first piece of research started before Pearl Harbor when the Chemical Warfare Service asked me and Mr. V. J. Schaefer, who was working with me as assistant, to undertake a theoretical study of the fundamental nature of filtration in gas masks. Now gas masks use charcoal to absorb poison gases, but even in the First World War, the possibility arose that the enemy might use smokes of toxic qualities, something not very volatile. These would remain suspended in the air in the form of dust particles which are not gases and therefore cannot be absorbed by charcoal and so have to be removed by a filter somewhat like filter paper.

The question asked us was "what are the principles that are involved in building a filter that will do as good a job as possible?" We were likely to be caught by surprise if the enemy de-

* Address delivered before the National Academy of Sciences, November 17, 1947.

cided to use toxic smokes for which we had no adequate filters.

The first problem was to make some smokes that would contain particles of a size that are the most difficult to filter. We wanted them to be stable, we wanted to measure their concentrations, and we wanted to make measurements of their particle sizes. That led us to an investigation of these factors.

It then took a lot of preparation to make particles of uniform size, determine how to measure them, and to learn how much of the material went through the filter.

That work lasted for about one year. We obtained fairly successful theoretical results and a better understanding of how to build a good filter. But notice what we did incidentally: we acquired a great deal of detailed knowledge as to how to make a smoke which would be non-volatile, which would consist of very, very small particles, far smaller than those of any ordinary smokes, and we learned much about their optical properties. We studied such things as the passage of gases through small spaces and around fibers and studied the trajectories of little particles carried by those gases.

This was done through an NDRC (National Defense Research Committee) contract. As we were getting toward the end of that work, a form letter was sent out (in August 1941) probably to everybody working in that field saying that the Germans were using on a somewhat large scale smokes for screening industrial plants. They were light colored smokes. They had only been seen from the air. Nobody knew what they were, but they were fairly effective. The question raised was, "could anyone think of a way to make a white screening smoke that could be used over large areas to cut down the hazard from aerial bombardment?"

Mr. Schaefer and I talked it over and we wondered whether we could not use the methods we had adopted for making smokes for testing filters. We had found that the easiest way to make smokes and control the particle size was to take some oil and put it into a volatile condition. We heated oleic acid and similar substances up to about 200 degrees centigrade and passed a stream of air over it so as to get its vapor mixed with air and then we quenched it suddenly by blowing in a large amount of cold air. The particles grew in size, and by sudden quenching we found it possible to stop the growth of the particles at a certain point and thus make particles of very small size. We were much surprised to find that under proper conditions we could get particles of extraordinarily uniform size.

Now the question confronting us was, how to do a similar operation on a large scale. We would have to use a material of low volatility, a petroleum product, perhaps, a high-boiling-point-lubricating oil or something of that kind, raising it to high temperatures.

From our optical studies, we knew that the small particles were the ones that did the best screening job. They scatter more completely and more effectively than the larger ones. It was like the scattering of light from the air molecules which produce the blue color of the sky. According to Rayleigh's law, such particles scatter with the sixth power of the radius R. For larger particles, the scattering varies as R^2. It is then proportional to the intercepted area. Now if you divide R^6 and R^2 by R^3, which is proportional to the volume of a droplet, you get a measure of the effectiveness of a given amount of material in scattering the light. Thus for very small droplets the efficiency increases with R^3 but for larger particles it varies with $1/R$ or inversely as the particle radius. In other words there must be a definite particle diameter which gives the most effective screening. This optimum diameter was found to be practically equal to the wavelength of the light which it is desired to scatter, about 0.6 micron for visible light.

These were much smaller particles than had been made by other generators but were just the particles we had been using in our work on the filtration of smokes. The hardest particle to filter is about 0.5 micron in diameter, and the ones best for screening purposes were about 0.6 micron in diameter.

The principle we adopted was that, if we wanted to get small particles, we would have to quench the vapor very fast. It is rather difficult to do this on a large scale. Therefore, if we wanted to form smoke on a bigger scale, we had to use higher velocities. We therefore decided to look into the possibility of using the highest possible velocity, the velocity of sound. We would have oil vapor come out into the atmosphere at about the velocity of sound. We would generate the vapor in a boiler at about ten pounds per square inch and let it come out through a nozzle at a velocity of nearly 500 feet per second. The air which is near

the jet of vaporized oil mixes with the vapor very rapidly and produces a cone-shaped jet.

We made calculations which were fairly elementary but gave important results. The theory was merely this: Very small particles having a diameter of one micron or less have high internal pressures because of surface tension. A simple experiment to illustrate this effect may be made with two soap-bubble pipes. By blowing a little bubble on one pipe and a big one on the other and connecting them with a rubber tube, one sees that the little one always disappears and adds to the size of the big one. The pressure in the little bubble is always more than in the big one.

It is the same with a droplet. The surface tension always exerts a higher pressure per unit area in the smaller drop. That means there is a tendency for the liquid to evaporate from the little one to give vapor which condenses on the big one. That tendency can be calculated. It was done by Sir William Thomson years ago and the equation is known as the Thomson equation. With droplets of the size we were concerned with, the difference in vapor pressure amounted to a few hundredths of one per cent. We heard later that the English had been interested in this problem and had considered this mechanism for the growth of particles in smokes. They too were trying to make screening smokes, and had thought of the theory that the small particles might disappear, giving vapor which condenses on the larger ones. They applied Thomson's equation, found that the difference was only one hundredth of one per cent and said, "It is negligible; let us forget it and find another way of doing it."

Fortunately, we went another step and we said, "Even if there is only one-hundredth of one per cent difference, the temperature is high since the oil is heated to its boiling point so the actual increment in pressure may not be negligible. As it cools in a few thousandths of a second, there is a period when even one-hundredth of one per cent difference in pressure may still be sufficient to permit evaporation of the smaller droplets.

How long under such conditions will it take the small particle to evaporate, disappear, and condense over on the big ones? When we made the calculations, we found it would take only a few thousandths of a second, which is just about the time it would take to mix the vapor with the air.

So quantitatively the theory gave results in accord with our first observations.

We could thus calculate, with oil vapor at, say, 10 pounds per square inch escaping from a nozzle of any given size, how big the particles would be in the resultant smoke. We found that unless the nozzle was pretty small the particles would be too big and we finally concluded that using a 3/16 inch diameter, with pressure of 10 pounds per square inch, was the way to get 0.6 micron diameter particles. Today the Army still uses nozzles of that diameter although we calculated the dimensions before our experiments had progressed very far.

The theory was improved, but it was fundamentally the same theory we had used in connection with the smokes we had tried to make for gas mask filters. The work went ahead very rapidly. We first started with one nozzle with 10 pounds pressure. That would use about 10 gallons of oil per hour in operation. We made up a little boiler and boiled lubricating oil at about 452 degrees centigrade, and let it come out through the nozzle at sonic velocity, and we found we got the right size particles of remarkable uniformity. They were very effective as a screening smoke.

One of the first field generators we constructed is shown at *B* in figure 1. That was a little thing, crudely made. It boiled 10 gallons of oil per hour.

Now the Army sent us some of their smoke generators to test in comparison with our field generator. They used partial combustion of petroleum oil, making a sooty, black, greasy smoke. When the Army adopted our smoke generator, they first cancelled a very large order for these smoke generators.

This is another one of these "threads" of which I speak: In all this work, on the problems I tell you about tonight, Mr. Schaefer and I have been associated. Vincent Schaefer is a good deal younger than I am. He came to the laboratory, working in the machine shop, and was a very skillful mechanic. He made a lot of fine apparatus for us. Later, he wanted to do research work. He worked on surface chemistry problems with us, did excellently at it, has published work on it, and has been working with me on these war jobs as they came along. He likes mountain climbing and I like mountain climbing; we like outdoor life in general.

You will see that such things are very important, and that they are closely related to our subsequent interest in weather control. Many things we are now doing have originated from that common background. Nobody could have planned

that for us. If we had not been naturally inclined that way, no one could have prevailed on us to go up on Mount Washington in the winter-time to carry on the research that occurred to us as a natural development.

Vincent Schaefer knew about the Schoharie Valley. We wanted a big wind tunnel. I mean a "big" one; one mile wide and ten miles long, that's all. We said, "Let's go to the Schoharie Valley." There is a valley ten miles long, about a mile wide, with a flat bottom, a cliff on one side, high hills two thousand feet high on the other side, and this cliff, seven hundred feet high, overhangs the valley right where the valley bends so you can look in both directions. We decided to make smoke in that valley and see how long it lasted and how far it traveled.

Figure 1 shows at *B* a line of smoke which goes all the way across the photograph. That is our generator, boiling ten gallons an hour. At *A* is the Army smoke pot that burns the same amount of oil, ten gallons per hour. That is what the Army was depending on for producing a screening smoke.

When we measured the relative performance of these units on other photographs, we found the area covered was four hundred to one in favor of our generator over the best one the Army had available. Recently at a meeting in Edgewood I found that they used to have a unit of 100 yards square for the old smoke pot in measuring the screening areas. After they got the new generator, they started to figure in terms of square miles. A mile is about 20 times 100 yards, so the new unit of area was 400 times the old one—close to the value we obtained in our first tests.

The improvement resulted from the use of the proper particle size, and from using a less volatile oil. The Army smoke pot depended largely on the soot produced since the oil was so volatile that most of it evaporated.

FIG. 1. A comparison of two smoke generators, Schoharie Valley, April 22, 1942. *A*. Army M-1 smoke pot. *B*. General Electric Research Laboratory smoke generator. Both units are using same amounts of smoke-making liquid.

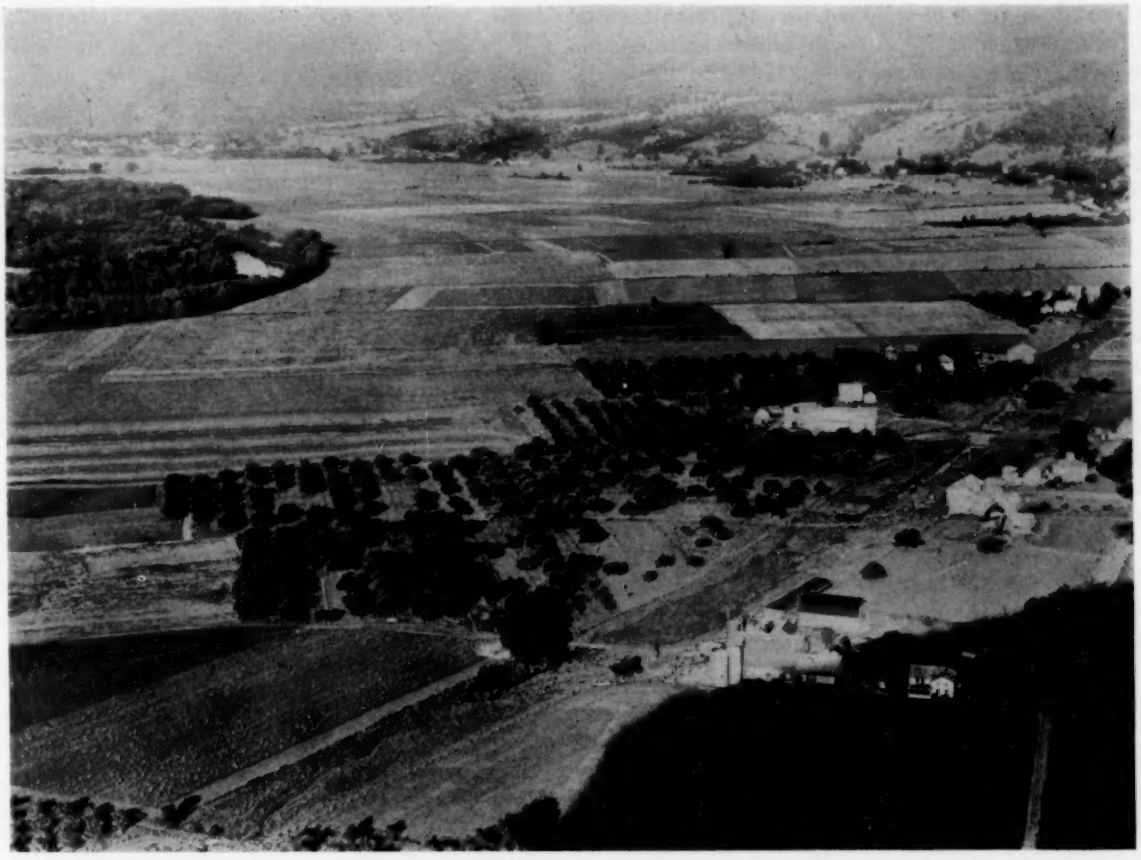

FIG. 2. Schoharie Valley from Vrooman's Nose.

The photograph of the Schoharie Valley in figure 2 shows the view to the southwest looking down from the top of Vrooman's Nose from an elevation 700 feet above the valley floor. The first tests were made at the base of the hill, and further tests with a larger generator which used 100 gallons per hour was up at the right hand side, near the top of the picture.

Figure 3 shows the same valley with one of our test clouds of screening smoke. There is not much wind blowing, but it shows the kind of job which can be done with a single generator.

When we made our final tests with Army and Navy officials present, it was a rather interesting day. So much depended on the weather to show the various properties of our screening smokes. We had about 30 to 40 Army and Navy officials, whom we had to get up at 2:30 in the morning. According to our prearranged schedule, we had breakfast at 3:30 and started off in jeeps long before daybreak. This was on June 13, 1942. We had to get them to the top of this mountain with only a tote road for the jeeps and then on foot paths. We all arrived half an hour before sunrise.

There were no casualties. The sun rose, there was not a cloud in the sky, there was very little wind, and the valley was completely filled with smoke in a short time. After we had run for about one hour and the whole valley to a distance of six or eight miles was filled with smoke, the wind started up as we hoped it would, and it blew the smoke out, to show what happens when you do not have quiet conditions and the smoke is carried into the sky. We were glad that this happened since it was important to show how an allowance must be made for the effects of winds when screening operations are carried out.

Figure 4 shows the large smoke generator used on June 13. The development of this generator is an interesting story. As soon as we demonstrated to our own satisfaction the operation

of our little ten-gallon generator, representatives of the NDRC and a few Army officers decided to embark on an extended program. The General Electric Company had no equipment for building the kind of apparatus needed for boiling of oils of the types we wanted. We went to the Standard Oil Company of New Jersey, and they built a large workable unit which was finished and ready for operation within a month after they visited us at Schenectady. We told them we wanted to boil 100 gallons of oil per hour, at 450° C.; and wanted it sent out through ten nozzles at 10 pounds per square inch pressure. This photograph was made a few hours after the first tests before the Army and Navy officers. This was 8:00 or 9:00 o'clock in the morning, after they had seen the valley filled with smoke from the mountain.

The generator was assembled by the Standard Oil personnel from things they could manage to obtain in three weeks from their scrap piles and other places. Other companies then started building them until they built many thousands of them. They were then used overseas and in the Pacific on a rather large scale. I have some pictures supplied by the Army, showing some of these field applications.

The city of Palermo, in Sicily, is shown in figure 5. You will see a line starting in the middle of the picture, going over to the left around some city blocks over there, where the smoke generators have just started up. Along that line, which includes most of the shipping, they had just started with the smoke generators. Just a few minutes later, it looked like figure 6. You can see the entire central part of the city was blotted out by

FIG. 3. Photograph taken from the same place as figure 2 but after the smoke generator had been put in operation and the smoke was approaching the point of observation.

FIG. 4. Large smoke generator used June 13, 1942, Schoharie Valley.

smoke. Later, they realized that better screening smokes from such generators could be produced by taking them further away and letting the smoke come in at a higher level. The smoke then did not enter low enough to interfere to any extent with operations on the ground, but planes above could not see anything on the ground.

Figure 7 shows one of the Army generators in actual service. Again, you see ten nozzles in operation. The generator there is more compact than the first one, and the later one is still smaller.

Since the generator uses 100 gallons of oil per hour, running continually, it posed a big supply job. I think altogether several millions of gallons were used. The Canadians used an enormous amount of it in Holland. Smoke generators of this type were used on a 65-mile front for three consecutive days before and during the crossing of the Rhine.

In the Pacific, the Navy first was not prepared to use the generators. They did not think they needed them, but after the kamikazi attacks were so successful they did decide to use them and as soon as they got them in operation, there was not a single ship damaged by kamikazi attack.

Figure 8 shows a close-up of the smoke coming out at high pressure from the ten nozzles. It is interesting to note that we can calculate that, in each second, a generator like this one produces 10^{17}, or 100 quadrillions of smoke particles. It does that so quickly and spreads them out into the air so quickly that there is no coagulation. The particles never get a chance to come in con-

FIG. 5. Palermo, Sicily, during the war, just after a set of smoke generators had been set in operation. (Signal Corps photo.)

FIG. 6. Same view as figure 5 a few minutes later when the smoke was beginning to screen the city. (Signal Corps photo.)

tact with one another in the process once they reach the optimum size.

That is an important thing. If you had that many smoke particles all from one nozzle, or if you had them in a confined space, the particles would grow to much larger size by coalescing with one another and the effectiveness would be ever so much less than it is.

I wanted to mention that because it is going to be possible to generate the same number of particles of silver iodide for their effect on clouds, when we get around to it. It is going to be possible to introduce from one generator about 10^{17} particles per second.

FIG. 7. The "M-1 Mechanical Smoke Generator" at the actual front after four hours of shelling by German artillery. (Chemical Warfare Service photo.)

Figure 9 shows another example of a smoke screen in use under combat conditions. It is a pontoon bridge crossing the Rhine, but it is obviously completely hidden from the air while the men continue to work on it. There is no appreciable smell, and no toxic effect from the smoke.

As soon as the Army assumed responsibility for obtaining generators, we dropped out of the field entirely. In fact, we did not know anything about what was happening with it until long after the war was over.

Again, quite independently of this, without anything originating with the NDRC and without any consideration of the fact that we had worked on a smoke generator, we were asked by the Secretary of War to look into the problems of precipitation static. This was in 1943.

It was believed that the invasion by Japan would have to come very largely from air attacks through the Aleutian Islands, across Alaska and from the North. That led to a tremendous development of air transport and airplanes through the Aleutians.

The difficulty in flying aircraft in the Aleutians is very, very serious. One of the great difficulties was icing of aircraft, but even more baffling was

the complete loss of radio contact when these planes flew through snow storms. In Alaska the planes become charged, sometimes, with 250,000 volts or more and this electric field produces corona discharges from all parts of the plane. This caused such electrical disturbances that radio sets could not be used to receive messages. Particularly, pilots had difficulty in finding their bases and getting down through this foggy bad weather.

Well, we had no particular ideas on the subject, except that it had to do with weather. It looked like a good opportunity to go up on Mount Washington in the winter-time. There could be no better place, we thought, to try something like that. Mount Washington in winter has a wonderful climate. The average temperature is 4 or 5 degrees below zero Fahrenheit. The wind velocity is about 60 miles per hour average, day and night, and most of the time with this high wind velocity clouds are sweeping over the summit.

We thought that, with this drifting snow, all we would have to do would be to take different kinds of surfaces such as aluminum, put different coatings on these surfaces, expose them to the drifting snow, and in the well-equipped laboratory of the Mount Washington Observatory, with power equipment and so on, measurements could be obtained of the electrical charges which develop from the impact of snow on the exposed surfaces.

We sent a lot of equipment up there before the road closed in the middle of October. We got it up there just in time. Then Mr. Schaefer went

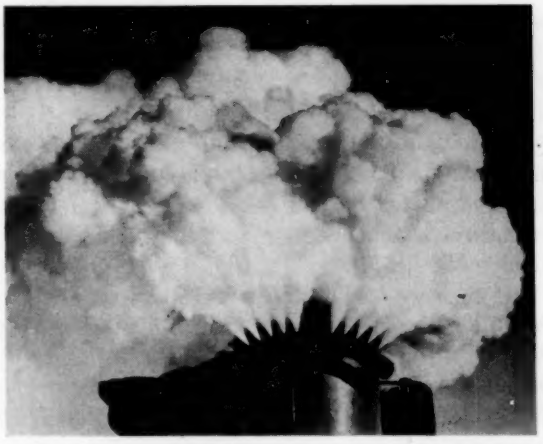

FIG. 8. Close-up of the ten nozzles of the M-1 "mechanical" smoke generator. (Chemical Corps School photo.)

FIG. 9. Trucks crossing the Rhine River over a pontoon bridge under cover of a smoke screen. (Signal Corps photo.)

up there several times to conduct experiments during storms and found the results to be very disappointing. We found, much to our surprise, that anything exposed on the summit of Mount Washington during the winter when clouds were there, as they are most of the time, immediately becomes covered with ice. This was due to the presence of supercooled water droplets. We had heard about rime, we had seen it, but we did not know it occurred all the time at a place like Mount Washington. There are very few winter days without it. This puzzled us a lot.

Now airplane pilots report that they have little, if any, icing on their wings and propellers when it is snowing. You know that when the air is above freezing, ice melts, and that below freezing water is supposed to freeze into ice. This means that ice is more stable. Once the water comes in contact with ice at temperatures below freezing, the water will turn into ice. That means ice has a lower vapor pressure than water. It has a lower tendency to evaporate. If you have a cloud which has water droplets in it that have not yet frozen although the temperature is below freezing, they will have a higher vapor pressure than ice. That is not uncommon, and I will return to this matter later.

I had carried out a lot of work on the evaporation of particles under the evaporation theory, determining how long it takes particles, small ones, to evaporate and condense on the large ones.

When you come to the difference between ice and water, there is, very often, 20 per cent difference in vapor pressure. It is 20 per cent instead of a few hundredths of one per cent as in the case of our screening smokes that I mentioned earlier. It is hundreds or thousands of times greater than in the other case. Therefore, the rate

of disappearance of the water droplets should be very large indeed in the presence of snow.

There is a limit, however. If you take a certain mass of air and lift it up into a cloud fast enough —whenever air rises into the atmosphere, the pressure becomes less. This lowering of pressure causes a lowering of temperature by the expansion of the air. The amount of cooling is about three degrees centigrade for every thousand feet you go up into the air, if the air is clear. If there is fog in the air, you get an added amount of moisture condensing out, so that some heat is generated by the condensation of that moisture. This decrease in temperature with altitude is less than three degrees within a cloud. As the cloud rises, water vapor is condensing all the time, and it tends to diffuse onto the snow flakes. It is a question, therefore, of whether the vapor pressure will get high enough to increase it as much as the 10 or 20 per cent necessary to give water droplets.

If the upward velocity of the air becomes high enough, water droplets may form. We made some calculations and compared them with the observed conditions on Mount Washington.

If you have a 60 mile an hour horizontal wind and there is a slope of 30 degrees, that means a 30 mile an hour vertical component of velocity, which is certainly a high velocity. With 30 mile an hour vertical velocities, you may generate water so fast by cooling the air that it does not have time to condense on the snow crystals and will actually form supercooled water droplets. That cloud, when it contacts a surface, deposits water droplets which immediately freeze on the ice already there. That is how ice builds up on aircraft wings.

We became so interested in this, that we hoped we would not have to continue a long study of precipitation static. It so happened that the Secretary of War's office was just as much interested in problems of aircraft icing as precipitation ice static. So we started on a study of the icing of aircraft. We had much assistance from the personnel of the Mount Washington Observatory. They were already working on riming and icing. They had cylinders and wires of different sizes which they would expose on the observatory tower. As the clouds passed these objects, measurements were made of the rate at which ice would accumulate on them. The cylinders would be exposed for fifteen minutes, rotating slowly so that ice would not form all on one side. They would then measure the diameters and the weights of accumulated ice and then put them back and run them for ten or fifteen minutes more, measuring the wind velocity at the station, and thus determine how much accumulated in a particular time interval.

They tried to interpret their data and calculate from it the amount of supercooled water in the air and the sizes of the particles. They obtained fairly good results for the amount of supercooled water in the air. However, with regard to their particle sizes, they had put too much dependence on certain photographic measurements which they made of slides exposed to the passing clouds. What really happened was that the droplets coalesced into larger droplets before they were photographed. They did not realize that the droplets photographed were not the primary ones at all, but the coalescence of small droplets into larger ones.

I found that the Germans and later on the English had worked on the theory of the deposition of ice from water droplets on rotating cylinders. The problem was one of trajectories.

Imagine a cylinder and the air going past it. Now the air does not go through the cylinder, it goes around it. It blows around it the way wind blows over an airplane wing. There is not much skin friction. The velocity is highest closest to the surface. When the velocity is high and the objects are large, the particles move with the wind around the cylinder. Larger particles of fog may not be able to follow that curvature. The particles act as projectiles. If the particles are big enough, there might be a 100 per cent efficiency of collection. That is, the cylinder would collect all the moisture in a volume of air equal to the projected area of the cylinder, multiplied by the velocity at which it is moving. That would be 100 per cent efficiency.

Now if, on the other hand, the particles are small or the wind velocities are low or the cylinder is large, many of the particles will follow the curvature of the cylinder and the collection efficiency will be low.

In order to determine the efficiency, it is necessary to make certain calculations just like those that the Ordnance people make in calculating the trajectory of a 16-inch shell in the presence of wind that may tend to blow it to one side. These are rather laborious calculations.

The original ones made were not applicable to such high velocities as those prevailing on Mount

Washington. We had to extend our calculations. We used for that purpose a differential analyzer which had been developed a few years before for making such computations much more easily. We had this machine working for about three weeks, working through the trajectories of particles passing over cylinders, and objects of different shapes, so that we could calculate the percentage of the water which would be deposited on a cylinder under specific conditions. We then hoped to be able to take this information to Mount Washington and so calculate the diameters of the particles and the number of the particles. We found it worked beautifully! We could take cylinders of different sizes and get values that agreed remarkably well with the theory.

We began to feel that we had here a very satisfactory method of understanding something about the structure of clouds and what makes particles grow.

At the base of a cloud on Mount Washington the air rises up into the cloud. It is forced to do so by the shape of the mountain. It is clear down in the valley, but as the air rises, it gets colder and colder and finally reaches the dew point and particles form. You may get thousands of particles per cubic centimeter, little droplets forming the cloud.

What makes them the size they are? "Well," I thought, "why isn't it the same thing as in a smoke generator? I will go back to the same evaporation-condensation theory that I used to calculate the rate of growth of smoke particles.

I found that the particles should decrease in number and increase in size, depending on the wind velocity and the liquid water content. We got values for the droplet diameter that agreed very well with those actually observed; in fact, the two sets of values showed a correlation coefficient of 0.94 which is, for meteorological problems, extremely high.

We then felt we had a fundamental theory for some of the factors that caused particles to grow in clouds to the size they are. However, the conditions on Mount Washington are complicated. The wind blowing over the irregular surface of the mountain gives turbulence and complicated wind structures. It would be far better to study cloud particle growth in airplane flights except that then, to measure the vertical air movements, new instruments would have to be devised. The velocities are higher and there are many other difficulties.

Schaefer and I gave a lot of thought to instruments that we could devise and put on airplanes to get the needed data. On a certain flight through a cloud, we would want to measure how much water there is, how big are the particles, and how they compare in size. What is the vertical velocity? How long has it been since that air entered the base of the cloud? How fast is it rising into the cloud?

Those were the things we wanted to measure. That meant an accelerometer or something to measure the bumpiness of the air. To get such data meant a lot of work.

The war had about ended at that time so we had no more Army contracts, or at least the war-time contracts were over, and I couldn't see why the General Electric Company should go into meteorological work. It seemed however that we should understand something about clouds from a national defense standpoint.

We took the question up with the Army Air Forces and the Signal Corps. We were led to think that perhaps somebody might furnish aircraft for experimental purposes of this sort. We did not get along very fast. We carried the research along on our own to a large extent, testing instruments on Mount Washington, but we never got tests in aircraft.

We were tremendously impressed by the relative scarcity of clouds on Mount Washington and elsewhere during the winter that contained ice crystals or snow. Ordinarily you cannot tell from the appearance of the clouds whether the temperature is above or below freezing. They look no different.

If a cloud contains snow, it does look different. We learned there were marked differences in the appearances of clouds. The disc of the sun as seen through a snow cloud is fuzzy. If the cloud contains water droplets, it is never fuzzy. We have made smokes of many droplet sizes, but the sun always showed a perfectly sharp edge as viewed through such smoke.

You can sometimes look at the moon and find the apparent disc is so big that it is hard to determine its size. Such effects are caused by cirrus clouds. There are lots of other things which characterize snow clouds.

The thing that struck us most is that, if there are any snow crystals in the cloud, they will be growing and falling. In the winter-time, if you see any stratus clouds from which no snow is falling, even though the temperatures in the clouds

are below freezing, there just simply are no crystals there in any reasonable number. Such clouds consist of water droplets. They can be supercooled to very low temperatures.

We thought that this was a problem that should be investigated. Why is it that sometimes snow forms so easily, with apparently no lack of ice nuclei on which crystals can grow, and at other times there seem to be none? There must be something in the atmosphere that causes water droplets to change to ice only at certain times and under various conditions.

Schaefer and I decided we must make some careful experiments in the laboratory to try to duplicate those conditions. I then went to California for three or four months. When I came back, Schaefer, I found, had made some beautiful experiments. He had taken a home freezing unit, which is used for food storage, about four cubic feet capacity and lined it with black velvet, directed a beam of light down into it, in order to see what happened in the chamber. He then breathed into it and the moisture condensed and formed fog particles which were just like ordinary cloud particles, although the temperature was about $-23°$ C. No ice crystals formed. He tried many different substances dusted into that box to get ice crystals to form, and practically never got any. He got just enough to convince him that, if he did get them, he could easily see them. However, the number was totally insignificant.

Finally, one day the temperature of the chamber was not low enough. He wanted it somewhat lower so he took a big piece of dry ice and put it in the chamber to lower the temperature. In an instant, the air was just full of ice crystals, millions of them.

He then took the dry ice out, and the ice crystals persisted for a while. Then he found that even the tiniest piece of dry ice would fill the cloud with crystals. Then he took a needle which he dipped into liquid air and he let that pass once across the box. The result again was the production of hundreds of millions of ice crystals by just one brief contact with the needle. The effect rapidly spread throughout the entire box. It is a wonderful experiment. You should look at it. It is very easy to do. It has been done by some high schools. One young girl student in Rhode Island made a chamber out of wash boilers cooled with brine. The effects to be seen are wonderful to look at and it is a simple matter to duplicate all the natural conditions of an actual cloud in the sky.

Well, that discovery changed the whole situation. What was discovered first was that the dry ice had no direct effect on the supercooled cloud but rather its temperature was the important thing. The needle, instead of being cooled in liquid air, could be cooled with dry ice. Instead of cooling it with dry ice, anything can be used having a temperature less than $-40°$ C.

Finally further study showed that there is a quite critical temperate of about $-39°$ C., the temperature of freezing of mercury, where a spontaneous reaction occurs to produce natural nuclei. This can be proved quite conclusively by covering the box with a sheet of glass and continuing to supply moisture into a cold chamber having a temperature of $-40°$ C. The ice crystals continue to separate out, fall on the bottom of the chamber, and new crystals form from moisture produced by evaporation of ice.

If, on the other hand, you take the box and cool it to $-20°$ C. and let air in from the laboratory, sometimes you find a few crystals, but if a cover is applied, they grow and snow out and there will be no more, unless you stir in some new air.

A little later, Dr. Vonnegut in our laboratory examined published data on the crystal forms of a whole series of other substances and he found that silver iodide has almost exactly the same crystalline form as ice and that it can serve as a nucleus for the growth of ice crystals.

You can make this experiment: Take a silver coin out of your pocket and while you hold it in your hand make a high frequency spark from a Tesla coil jump to it. At the moment the spark passes, you blow once down into the box which is open on top and full of cold air. Then a moment later you open a bottle of iodine and blow some of its vapor into the box. You soon find that the air becomes full of ice crystals, perhaps less than a millimeter apart, which means that something like 100 million snow crystals have been formed in 100 liters of air, from the silver iodide particles that were caused by the vaporization of the silver coin as you held it in your hand.

Electron microscopic pictures of these particles indicate that from one gram of silver iodide about 10^{17} nuclei can be produced. This suggests that a few pounds of silver iodide would be enough to nucleate all the air of the United States at one time, so it would contain one particle per cubic

inch, which is far more than the number of ice nuclei which occur normally under natural conditions.

That smoke generator I told you about earlier can generate 10^{17} smoke particles per second. If you make silver iodide smoke that way, it would cost five dollars an hour for the materials. The problem of getting a proper distribution of particles is more difficult. If the particles retained the activity that they have in laboratory tests a wide distribution of them in the atmosphere might perhaps have a profound effect upon the climate.

Table 1 shows the relationship of temperature and vapor pressures. The first column gives the temperature. The 2nd and 3rd columns give e_w and e_s, the equilibrium vapor pressures for water and ice or snow, expressed in millibars. The next column gives the difference between the vapor pressure of water and of ice. This reaches a maximum of 0.28 mb at − 12° C. The next to the last column gives this difference in vapor pressure expressed as a fraction of the vapor pressure of water. Thus at − 40° C., the vapor pressure of water is 33 per cent higher than that of ice.

TABLE 1

VAPOR PRESSURES OF WATER AND ICE AND THE CRITICAL DIAMETERS OF ICE NUCLEI

Temp. °C	e_w mb	e_s mb	$e_w - e_s$ mb	$\frac{e_w - e_s}{e_w}$	Crit. diam. cm.
0	6.11	6.11	0	0	∞
− 1	5.27	5.17	0.10	.019	7.8×10^{-6}
− 5	4.19	3.99	0.20	.048	3.10
−10	2.86	2.60	0.26	.091	1.68
−15	1.89	1.64	0.25	.133	1.18
−20	1.25	1.03	0.22	.174	0.93
−30	0.51	0.38	0.13	.256	0.67
−40	0.19	0.13	0.064	.332	0.56

The last column shows how large an ice crystal would have to be if it is to grow. Very small ice particles would have higher vapor pressures than larger ice crystals and, if they have diameters less than the critical diameters shown in this table, they would have vapor pressures greater than that of *water* and therefore would not be able to grow even in a cloud of supercooled water droplets. It appears, however, that in supercooled clouds at temperatures below − 5° C. all ice particles (nuclei) greater than 3×10^{-6} cm in diameter should grow.

Table 2 shows the distances that dry ice pellets should fall in free air before they completely evaporate. Thus a particle 0.4 centimeter in diameter would fall 1,000 meters before it disappeared. These calculations gave us some idea of the size pellets we would need, in order to have them go down through a cloud.

TABLE 2

THE FALL OF CO_2 PELLETS THROUGH THE AIR AT − 20° C.

Pellet diameter, cm	Velocity, cm/sec	Time to end of fall, sec	Distance to end of fall, meters
1	2120	350	4,300
0.4	1420	127	1,130
0.2	950	59	330
0.10	564	26	82
0.04	233	8	10
0.02	105	3	1.6
0.01	41	1	0.3

Using principles which I developed during our studies of the growth of smoke particles and the deposition of rime I have been able to calculate the amount of water which is condensed from the vapor by contact with the falling pellets of dry ice. The results are given in the 2nd column of table 3, while the 4th column gives the volume of air cooled per second to a temperature below − 40° C. at which ice crystals form spontaneously. The concentration of nuclei formed initially is so great that even during the short time of contact with the pellet the particles grow by coagulation giving the concentration shown in the 3rd column. From these data it was then possible to calculate the diameters of the nuclei (5th column) and the total number produced (6th column). The experiments seem to indicate the calculations are fairly accurate. The most minute particles of dry ice dropped into Schaefer's cold box produce hundreds of millions of ice crystals.

November 13, 1946, Schaefer dropped about six pounds of granulated dry ice into a supercooled strato cumulus cloud at the rate of about a pound and a half per mile. He then flew back to watch the results and they were very striking. The whole supercooled cloud was converted to small ice crystals.

Since then, starting in March, we have had a contract with the Signal Corps. This work, Project Cirrus, is being done by the Signal Corps and the Navy with the close cooperation of the Air Forces. The operations are all carried on by an

TABLE 3

NUMBER AND SIZES OF ICE NUCLEI PRODUCED IN CLOUDS AT − 20° C.

(1) Pellet diameter, cm	(2) Total wt. of H_2O condensed in remainder of fall, milligrams	(3) Concentration of nuclei/cm³	(4) Volume of air cooled, cm³/sec	(5) Diameter of nuclei generated cm	(6) Tot. no. of nuclei produced in remainder of fall
1	960	3×10^{13}	70	2.9×10^{-6}	1.7×10^{17}
0.4	61	5	18	2.2	2.6×10^{16}
0.2	7	7	6	1.9	6×10^{15}
0.1	1.0	8	2.2	1.6	1×10^{15}
0.04	0.06	8	0.5	1.4	1.2×10^{14}
0.02	0.007	7	0.2	1.4	5×10^{13}
0.01	0.001	56	0.1	0.6	2×10^{13}

Operations Group which consists wholly of men from the armed forces. Figure 10 shows a flight test made April 7, 1947, in which stratus clouds were seeded with dry ice. The tops of the clouds were at about 6,700 feet altitude. The cloud layer was about 4,000 feet thick. This picture was made about 4,500 feet above the top of the clouds. The long dark band seen in the picture is only part of the seeded region which is about 12 miles long, with a width of about 3 miles. This area consists of snow crystals while the rest is made up of supercooled droplets.

The shaded area looks dark on a white background. The ice crystals and water droplets scatter the rays of light in different ways. From certain directions an ice crystal cloud looks darker in color than the supercooled clouds, but in other positions with respect to the sun it is lighter than the surrounding cloud.

Figure 11 shows the same region at an earlier time as seen from a different direction. The seeded line was in the form of an "L," about 6 miles along each leg. Here, it looks like a white river.

Figure 12 is an interesting photograph. You see, directly below the sun, a specular reflection in the seeded area. That is a characteristic of ice crystals in the form of hexagonal plates. They tend to fall horizontally and, if you maneuver the plane so the sun is in the right position, a high intensity reflection is seen owing to the fact that you are looking down on these little flat crystals, which are floating horizontally. This "sun column" below the sun is conclusive proof of the existence of ice crystals. It is never produced by a supercooled cloud.

We have never found a case with temperatures below freezing where an ordinary cloud cannot be converted to ice crystals by the introduction of dry ice.

Figure 13 shows the distances in miles. That affected region covers an area about the size of Lake George in New York. It is several times as big in area as Manhattan Island. We purposely sowed a small area so it would not be too large to photograph.

This winter we hope to be making experiments with five Army and Navy planes. When I say "we," I mean the Operations Groups controlling those planes.

FIG. 10. A part of the L-shaped pattern seeded in a stratus cloud at 4:39 P.M. The airplane was at an altitude of 11,200 feet at 5:04 P.M. on April 7, 1947. (Signal Corps photo.)

FIG. 11. The same L taken at 4:54 P.M. from an altitude of 9,100 feet. (Signal Corps photo.)

We have not done extensive work on these stratus type clouds. It is very evident in the results so far that with stratus clouds and one airplane, something like 500 square miles of supercooled clouds can be converted to ice crystals per hour of flight. Up to the present we have not attempted any such large scale experiments for it would be difficult to keep track of the changes occurring in a 500-square-mile area. This winter, probably off the coasts, maybe in Alaska, we hope experiments will be made on a very large scale, to see whether or not it is possible to remove completely icing conditions for aircraft over big areas by converting all supercooled water clouds in the area to ice crystals, wherein there would be no trouble with icing.

In some of these flights last winter, on the ground, you could see, within ten to fifteen minutes after seeding, a great open space develop. I saw at one time, in Schenectady, fully 100 square miles of clear sky on a day when there were no openings whatever in the cloud before seeding, and none after that, except this one area about 20 miles long and 5 miles wide from which you could see snow falling from the edges with the whole thing drifting over toward the mountains. The results were interesting, although very hard to photograph because they were up so high.

We have also been seeding cumulus clouds. We have been working up techniques, training personnel and getting things really organized for effective work this winter. It is surprising to note that during July and August, in spite of the fact that generally there was very little rainfall, there were a surprising number of days where, from an altitude of 5,000 feet or more, several very large cumulus clouds could be seen from which no rain was falling. These clouds looked like thunderheads, but they had not built up to that stage. They often go to 25,000 and 30,000 feet. They were not all like that, but frequently they are up that high. When you seed such clouds, the general observation is that things happen very fast.

FIG. 12. A photograph of the L at 5:16 P.M. from an altitude of 12,000 feet facing in the direction of the sun.

FIG. 13. One end of the L taken at 5:24 P.M. and at an altitude of 14,800 feet. This shows the actual dimensions in miles calculated from the angles measured from the photographs.

The top of the cloud may be below freezing for only about 4,000 feet. The freezing level may be 18,000 feet and the top may be some 26,000 feet.

If dry ice is dropped anywhere in the top of such a cloud, the results are often unusual, to say the least. In one case, I happened to be driving an automobile under a cloud which was being seeded. I never saw such heavy rain in all my life. It came down in torrents for about fifteen minutes. The wind was so strong that it looked as though it was going to blow over several trees, and a few trees were blown over. Fortunately, it did no harm. We drove up to the side of the road, my wife and I, because we did not know what was going to happen. Then there was a little hail, small pieces about a half inch in diameter, after the rain. Usually, the hail comes first, but here it had been raining for five or ten minutes and then came some hail.

It then let up a little, so I drove on, less than half a mile. The rain stopped. I got out and looked, and the road was dry. It had not rained there at all.

We expect such thunderstorms to go to the east in our region. That storm did not go to the east, it just ended, completely. At that time, five minutes after it stopped raining, there was no rain to be seen anywhere. The whole thing disappeared. Now that is a quite unusual type of thunderstorm.

Other people have done the same thing in other parts of the country and although we have no idea how many are correct and not correct, the uniformity of results seems to be quite striking. The reports indicated that great effects are produced within ten or fifteen minutes. Rain usually reaches the ground heavily in fifteen to twenty minutes and usually the rainfall is over in a half hour. In a few cases, they claimed it rained for a long time—several hours—getting an inch or more of rain. An inch of rain seems to be very common over a fairly limited area. That is about what might be expected from the water content of such clouds.

In other words, it seems to precipitate most of the moisture within a very short time. It was a puzzling thing to me to know how things could go so fast; how the rain can get down so fast.

The effect of the dry ice is to cool a little of the air below $-40°$ C. and so produce great numbers of ice crystals which then proceed to grow until they have used up all the available water in the cloud. The larger the number of ice crystals introduced into a given volume of cloud the smaller will be the size of the snow crystals that are formed. Near the edges of the seeded area, there are rather few ice crystals so they grow to large size and fall rapidly. In spite of the fact that you introduce enormous numbers of ice crystals into a cloud, they will ultimately run out, unless something generates more.

In nature, there is no question that snow storms do continue for a long time, although there are few ice nuclei originally present in the air. There must be some process by which new nuclei or ice crystals are generated and that causes the development of a chain reaction. The chain reaction is one where you start out with one snow flake and, by some mechanism, it becomes two. Then those two become four, and those four, eight, and so on, and the thing propagates rapidly, in exactly the same way as does a haystack when you light it with a match. It starts burning. It does not make any difference where it is ignited, in three places or one place, it spreads through the whole haystack. The heat produced by burning one particle of hay has to be enough to heat the next particle.

So it is with snow flakes. Now I do not know, but I suspect that the chain reaction that goes on in nature's storms is a fragmentation of snow crystals. We made a study of these phenomena and we think that electrical effects take place later that stimulate the process. We can stimulate the growth of ice crystals after they are once formed by electrical means, but we cannot initiate them. We can pass large electrical charges through our box, but never produce any ice crystals in a supercooled cloud. If we put dry ice into a supercooled cloud in the cold box and apply a strong electric field the ice crystals grow into long needles or fibers which in a sufficiently strong field break apart into pairs of highly charged fragments. So we think that the electrical effects do play a role, but only in the later stages.

Now there is a chain reaction which we may be able to produce in winter snowstorms, by dry ice, if we seed storms that are naturally growing, where there is an ample supply of moisture. We have tried to avoid such things in populated districts. We do not want to have a snow storm over New York City which might cause New York City to spend $5,000,000 to gather up the snow. It is extremely important to find out if such things are possible, and to learn exactly how to get the best results, out over uninhabited regions. Until we have learned the techniques and have then carefully considered through the government agencies what the American public as a whole wants, we must proceed carefully.

Out in Hawaii, recently, there was an extremely interesting flight, by Luna B. Leopold and Maurice H. Halstead, where dry ice was dropped on some clouds at a temperature of $+10°$ C., and they got rain. The cloud started to grow from 8,000 feet up to 15,000 feet and then rose above the freezing level when it was again seeded with more dry ice. It grew to 26,000 feet and kept on raining and spread and traveled out across the sea and went from one island to the other and rained all night.

The results of the experiment made me think a lot, because I have been interested in the growth of rain ever since 1944 when I started to develop a mathematical theory of the growth of particles in clouds.

Now this summer, in the seeding of the cumulus clouds, only the top was above the freezing level and the snow that formed in that top came down and melted and made water droplets which then gathered in all the water in the cloud and dissipated the cloud, all within fifteen or twenty minutes. It was something that happened throughout the cloud, not only at the top of the cloud. This seemed to me to constitute a chain reaction.

I began to think more about the problem. I had the mathematics all worked out, but I had not thought of it as a chain reaction. I applied it to the reported cases in Honolulu. It seems to me that the effect is really a very simple thing and I wondered why no one had thought of it before. If, in a cloud, there is one large drop of water, just one, and it falls, it will grow as it goes down; that is, if it is larger than a certain size. To find out what that size must be, it is necessary to use mathematical methods such as we have used in calculating the formation of rime on cylinders. I have been making these calculations during the last three months, so I know just how big a drop must be before it will grow when it falls. If it grows, it grows exponentially. It grows faster the bigger it gets, and then it falls faster. As its cross section increases in area it gathers intercepted droplets more rapidly and it thus sweeps through a roughly conical shaped volume, increasing in size toward the base of the cloud. It grows faster and faster until it gets so big, finally, that it breaks up, producing smaller droplets.

Now if there are no rising air currents all those drops fall until they reach the bottom of the cloud. That is no chain reaction.

Suppose, however, there is an updraft so that the little droplets that break off when these drops get as big as they can get are carried aloft. By the time they come back to the level at which they started, they may have grown again to the point where they have to break up. Then one drop produces two or ten and each of those produces two or ten. As long as the factor is more than one, it is a chain reaction and the whole cloud simply goes over into heavy rain.

Now I believe that this has already been done, in the experiments in Hawaii. It was done by us every time we seeded a cumulus cloud. I have worked it out mathematically in a theory which will be published soon.[1] It is simple. It tells you that the conditions that have to be met in order to bring about that chain reaction in rain formation are a vertical velocity of five miles per hour and a liquid water content such that there is at least a tenth of an inch of rain available. The cloud particles have to be rather large, about twenty microns in diameter; very small ones will not do. When you meet those specified conditions then only a single large drop of water will set off a whole cumulus cloud and bring down most of the moisture in it.

Seeding with water will be just as good as dry ice and far more frequently applicable all through the tropics where low temperatures do not occur.

The literature is full of examples of people who have observed that sort of thing and never understood it. Airmen all through the war have seen heavy tropical showers but have not understood how such heavy rain could occur without snowflakes near the top of the clouds to start the growth of the raindrops.

Of course, we hope to try this type of cloud seeding on a large scale in the tropical islands to see whether water seeding does not take its place along with dry ice.

We have done two other things in which you will be interested. One is the seeding of New England clouds to try to produce rain enough to discourage forest fires. The result is inconclusive, because scattered showers began to form that day, starting in about one or two hours before we seeded.

Now the other thing of interest is our activity with regard to hurricanes. You will remember there was a hurricane on September 23 that struck Miami. The newspapers were full of accounts that certain individuals were planning to seed it. They called it "hurricane busting." I do not know where that phrase came from, because nobody intended to bust the hurricane, as far as I know. A flight was planned to get information. However, we were not ready for that. It takes much work and planning. Lieutenant Commander Daniel Rex has done a good job in organizing the Army, Navy, and Air Forces groups participating in Project Cirrus. The commander of this project is a Naval officer, though most of the planes are Air Force planes. There is really close harmony and it is a wonderful team. However, they were not ready for a large operation such as studying the possible effects of dry ice seeding upon a hurricane. They did not have the equipment on the planes and the men had not been properly trained.

The next hurricane came, or we first heard of it, on October 10. The planes were ready to leave on October 11. They flew down to Mobile and then to Tampa to make the study. The next day, Sunday, the hurricane had moved with extraordinary speed. It crossed Cuba and went suddenly into Florida and was soon out at sea, beyond Miami.

[1] This chain reaction theory was presented at a meeting of the American Meteorological Society in New York January 28, 1948 and will be published in full in the *Journal* of the society.

They could have gone out on Sunday afternoon and tried something but decided not to do that since the observation conditions would not be ideal.

The hurricane was actually seeded experimentally on October 13. These experiments were made under the control of the Operations Group of Project Cirrus. I was not present at any of them. Mr. Schaefer and one General Electric man were each on one of two planes as observers.

The following is a report that was prepared by Commander Daniel F. Rex, Chairman of the Operations Group:

On 13 October, 1947, a small scale seeding of a tropical hurricane located 350 miles east northeast of Jacksonville, Florida, was effected by the Operations Group of Project CIRRUS. The group, consisting of three aircraft, was based at Tampa (MacDill Field) Florida. Takeoff was effected at 0820 EST Monday morning. The group proceeded in company to the storm area arriving at the outer wall of the eye at 1045 EST.

The storm consisted of an eye approximately 30 miles in diameter surrounded by a thick wall of clouds extending from about 800 feet up into the cirrus overcast at 20,000 feet and being some 30–50 miles thick radially. Several decks (4 or 5) of stratified shelf clouds extended out from the outer wall, the uppermost deck having tops at 19,000 feet. These shelf clouds appeared as large areas (100–200 square miles) of solid, thin (1,000–2,000 feet thick) undercast separated by large breaks through which the surface was often visible. An exceedingly active squall line, appearing as an almost continuous line of cumulonimbus with cirrus tops to an estimated 60,000 feet, was observed as a spiral extending out from the center to the east and south. A well developed cirrus dome (funnel) extended out from the center-base at 20,000 feet near outer wall lifting to 35,000 feet at edge.

Approach to the storm center was effected from the southwest, this course bringing the group into the storm's right rear quadrant. After a brief reconnaissance flight around the outer wall, the decision was made to seed a track over the uppermost cloud shelf and at a distance from the center sufficient to permit the control aircraft to fly contact 5,000 feet above the seeding aircraft. A formation intrail was used with the seeding aircraft (B–17 #5560) leading at cloud top level. The photo-reconnaissance aircraft (B–17 #7746) followed the seed ship, 3,000 feet above and 1–2 miles astern with the control aircraft (B–29 #816) trailing 5,000 feet above and 15–20 miles astern. Seeding commenced at 29.8 degrees North, 74.9 degrees West at 1138 EST at an altitude of 19,200 feet, the outside air temperature being approximately −5 degrees centigrade. Continuous seeding was effected along a straight course to 30.2 degrees North, 73.9 degrees West, thence to 30.8 degrees North, 73.1 degrees West at which point (1208 EST) seeding was stopped. During this thirty-minute period 80 pounds of solid carbon dioxide was dispensed along the 110-mile track. In addition, two mass drops of 50 pounds each were made into a large cumulus top at 30.7 degrees North, 73.4 degrees West. Upon completion of this phase all planes flew a reverse course back along the seeded track taking visual and photographic observations. No attempt was made to penetrate through the wall of the storm into the eye or to seed in or near the above mentioned squall line due to the failure of the group's homing aids (radio compass and visual flares). It was thought that such an attempt, although desirable, would likely result in a separation of the aircraft with subsequent abortion of the primary mission.

Visual observation of the seeded area showed a pronounced modification of the cloud deck seeded. No organized trough was observed; rather the overcast previously observed appeared as an area of widely scattered snow clouds. The disturbed area covered perhaps 300 square miles. No convective activity was seen to follow the seeding process at any time during the mission.

The group returned to MacDill Field, landing at 1600 on 13 October 1947. No post test observation flights were made.

Now the present hurricane started near an island west of Jamaica. It went north, followed across Florida past Miami, then on out to sea, and was 350 miles off to sea at the time of seeding. The circle shows the approximate area of the clouds of the hurricane. The seeding run is the little line running almost straight, just below the center of the circle as shown in figure 14. The squall line is that big arrow off to the right. The path of the hurricane from that point on was uncertain for about a day or so. The weather forecasters thought it had, at first, two centers, but later on decided it did not have two centers. The storm recurved and went west.

The extraordinary thing noticed about the storm by those on the plane was the extreme mildness of it. It was no storm at all in much of the area but seemed more like relatively fair weather. There were no cumulus clouds, there was no rain, the wind velocities were about 30 miles per hour, at the height of the flight, and the effect of seeding was what you would expect to get in similar conditions in Schenectady.

They felt however that if they could have been on the other side of the squall line, which I already mentioned, some really important things might

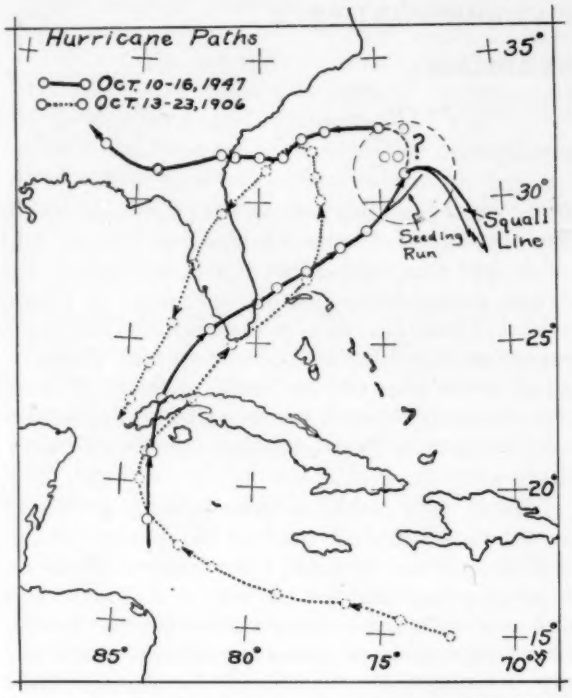

FIG. 14. The paths of the hurricanes of October 10–16, 1947 and of October 13–23, 1906.

have happened. What did happen, nobody knows. The squall line is not a formal feature of a hurricane. When it started, when it ended, what happened to it, nobody knows. The Weather Bureau has no record whatever. The squall line shows up on the photographs they took from the planes but apart from this there are no meteorological data that show the existence of the squall line, what it meant or what caused it.

The main thing that we learn from this flight is that we need to know enormously more than we do at present about hurricanes.

Consider, for example, the hurricane of September 23. It was far out at sea east of Florida and it started to move to the north or northeast and, according to the weather forecasts, it was expected to continue moving northeast parallel to the coast. Instead of that it suddenly turned around, went west and struck Miami.

In 1906 between October 13 and October 23 there was a hurricane which followed a path, shown as a dotted line in figure 14, very much like that of the hurricane of October 1947. It too turned almost back on its track and moved back to the coast. Perhaps it also had a squall line and this contributed to its erratic behavior.

We do know if we investigate these things long enough, we may be able to get an understanding of hurricanes.

It seems to me that next year's program should be to study hurricanes away from land, maybe out considerably beyond Bermuda, out in the middle of the Atlantic, make an extensive study of them, flying from the right to the left, forward and backward, and learning what happens, and see if we cannot, by seeding them, in some way modify them or shift their positions. I think the chances are excellent that, with increased knowledge, something can be done. The stakes are large and, with increased knowledge, I think we should be able to abolish the evil effects of these hurricanes.

THE FRANKLIN-GREENE CORRESPONDENCE

WILLIAM GREENE ROELKER

Director, Rhode Island Historical Society

(*Read November 21, 1947*)

THERE are many books about Benjamin Franklin; he has been pictured as the statesman and politician, the inventor and scientist, the philosopher and sage. His writings and sayings, his appearance and manner, have been described in many of the languages of the civilized world. But almost nothing has been written of the friendship between Franklin and Mrs. Catharine (Ray) Greene, which endured from the time they met at Boston in 1754 until his death in 1790. The correspondence with Catharine Ray, which began in 1755 and continued with some lapses until 1789, preserves the record of their long friendship. Yet in all those years they saw each other on but five occasions—for the last time at Philadelphia in 1776.

In all, about one hundred letters have been discovered, including the letters which passed between Franklin and Catharine, their several letters to his sister, Mrs. Jane Mecom, and her replies. A number of Franklin's letters to the two women have been printed with his other correspondence by Sparks, Duane, Bigelow, Smyth, and Van Doren. No attempt, however, has been made to bring them together, nor to integrate them with the many letters hitherto unpublished written to him by Catharine and Jane.

Catharine Ray is the figure around whom the correspondence centered. She was married, baptized, and buried as Catharine, although nearly all writers have persistently written Catherine. Block Island, the tiny islet where she was born and raised, is twelve miles out at sea from Point Judith, the extreme southernmost point of Rhode Island. It must have been a dreary abode for an intelligent, fun-loving, and attractive girl of twenty-three! All three of her sisters were married, and she lived alone with her aged parents. Her father, Simon Ray (an old New England friend of Franklin),[1] was then eighty-two and her mother sixty-five. The youngest sister, Phebe, was the wife of John Littlefield, an Islander; the other sisters had married on the mainland: Judith was the wife of Thomas Hubbard of Boston; and Anna had married Samuel Ward, son of Richard Ward, former Governor of the Colony of Rhode Island. Franklin was well acquainted with the numerous Ward family of Newport and Westerly, all of whom were tied by family, political, or business associations with the large body of descendants of Surgeon John Greene, a founder of Providence and Warwick.

"Katy" as Franklin addressed her, or "Caty" as she signed herself, married her second cousin, William Greene, in 1758. Her mother, Deborah, a great granddaughter of Roger Williams, was also a member of the extensive Greene family. Other Greenes who enter prominently into the correspondence are Nathanael, the Revolutionary General who married Caty's niece, "Kitty" Littlefield; and his brother, Elihu, who married Franklin's grandniece, Jane Flagg. Nearly all of these characters lived in close proximity.

On the Franklin side, after BF himself, by far the most prominent is Jane, his favorite sister and wife of Edward Mecom of Boston. Jane early became attached to Caty, whom she probably met at the Hubbard's after John Franklin had married the widow Hubbard in 1753. Caty and Jane were some twenty years apart in age, and their relationship was in a way that of parent and child, Caty calling her "my mama and friend." When Boston was under siege by the American army in the spring of 1775, Jane Mecom fled the town and took refuge with her friend, Mrs. Catharine Greene, at Warwick, Rhode Island. She continued a guest at Warwick until late in October, when her brother called for her and in his carriage escorted her, together with Caty's ten-year-old son Ray, to Philadelphia. The correspondence between Jane and Caty began at this time and continued until Franklin's death. Each of the women idolized Franklin, and there is never a letter from either one that does not inquire the latest news of their "dear Friend." In later years when Franklin was at Paris his every letter provided "a feast" for the two devoted women.

[1] Parton, James, *Life and times of Benjamin Franklin* I: 342, Boston, Ticknor and Fields, 1867

Another bond centered around Sarah Franklin, BF's daughter, who had married Richard Bache in 1767. Sally had visited Catharine in 1763 when her father was making a tour of inspection of the post offices; and Catharine had met Sally's huband and their children when she accompanied her husband to Philadelphia on Colony business in July, 1776. Because Jane and Caty both thrived on domestic history, as did BF himself, all their letters teem with the small details of family news. Thus around Catharine Greene at Warwick, the Hubbards at Boston, the Wards at Westerly and Newport, and the Franklin-Baches at Philadelphia there was a web of intimacy, friendship, and love which is reflected in this collection of letters. They portray the kindly, mellow, domestic side of Franklin as does no other group of the thousands of letters that he wrote.

Franklin had set out from Philadelphia in August, 1754, on a tour of inspection and regulation of the post offices. During November and December he seems to have been in the vicinity of Boston, where he had various conferences with Governor Shirley regarding the aftermath of the Albany Convention. But he did not neglect post office business. He had appointed his brother, John, a tallow chandler, in place of Elias Huske. The Boston *Weekly News-Letter* for January 2, 1755, had carried the following announcement: "Notice is hereby given, That the Post Office is now kept at Mr. John Franklin's in Cornhill, December 30, 1754."

Because in the eighteenth century it was customary for visitors to stay with their relatives, Franklin naturally put up at his brother John's, where Caty was visiting in the early winter of 1754–1755.

"Mr. John Franklin's in Cornhill" was the mansion which Mrs. Elizabeth (Gooch) Hubbard had inherited in 1734 on the death of her husband, John Hubbard. John Franklin came to live there when he married the widow Hubbard in 1753. From the Land Title Records in the Bowditch Collection at the Massachusetts Historical Society the exact location of this property has been established; the site is now occupied by the northeasterly part of the Ames Building. The Hubbard property fronted 21 feet on Cornhill (now Washington Street) and extended 144 feet towards Dassetts Alley (now Franklin Avenue). At the corner of the latter and Queen Street (now Court Street) was the printing office where James and Benjamin Franklin published the New England *Courant*.

About ten minutes' walk from the Hubbard mansion on Cornhill, at 19 Unity Street, directly behind the Old North Church, lived Jane Mecom. The house in which she lived was the only one in Boston that Franklin ever owned. He had acquired the property by foreclosure from his elder half-sister, Elizabeth (Franklin) Dowse, and her husband, Richard. After the death in 1753 of Captain Dowse, Edward and Jane Mecom with their many children came to live with the widow. When she died in 1759, Franklin deeded the house to Jane, and she continued to live there (except when she was with her granddaughter in Rhode Island in 1777–1783) until her death in 1794. In 1939 the house was purchased by the White Fund and torn down in order to extend the vista from the Paul Revere Mall to the Old North Church.[2]

Though the Hubbard homestead was called a mansion, it was only a twenty-one foot house, probably two stories high. At least two families lived there continuously, and there were many children. Because eighteenth-century business was usually conducted in the home, the street floor front was almost surely a store like the one later occupied by Paul Revere at 50 Cornhill nearby. In these crowded conditions Caty and Franklin must have been thrown together much of the time during their visits. What could be more thrilling than meeting the famous Dr. Franklin! It is true that he was nearly forty-eight years old, married, and had a twelve year old daughter. But he was a distinguished person and a charmer. It is not surprising that Caty was impressed, and the more so because from the outset he seems to have been greatly taken by her wit and gaity. Heady wine for a young country girl! She was fascinated and flattered by the attention he paid her. He thought "Sugar Plums" sweeter if made by her hands. He admired her mind and was stimulated by the quick responsiveness of her intellect. His ability to read her thoughts caused her playfully to accuse him of being a "conjurer." Twenty-five years later she wrote him at Paris: "I with Pleasure look Back Upon those Pleasant days I used to See you in Boston and the AGreeable Journey to Westerly which you have Some times Mentioned. . . ." [Dec. 29, 1780.]

Franklin and Caty had set out in his carriage from Boston on December 30, 1754, for Newport,

[2] Brown, Frank Chouteau, The Clough-Langdon House, 21 Unity Street, Boston, *Old-Time New England*, **37** (4): 83, 1947.

Rhode Island, where she had many Ward connections. No details of this memorable journey are known, but it may be supposed that they followed the usual route through Dedham, Walpole, Wrentham, Attleborough, Rehoboth, Warren, and over the ferry at Bristol to Newport. The distance was about seventy miles. If the speed of their travel can be judged by that made by Mrs. Abigail Adams in 1789, Franklin and Caty must have been a full two days on the road.

Their probable over-night stop at a tavern has caused some raising of the eyebrows as to their relationship. It has been insinuated that Catharine was just another of Franklin's conquests. But it may be stated without fear of contradiction that their relationship was quite Platonic, no matter what may have been his intentions. The evidence is nothing less than a statement in his own hand. In a letter dated October 16, 1755, Franklin gives Catharine "some fatherly advice." He tells her to get a good husband: "You must practise *Addition* to your Husband's Estate, by Industry & Frugality; *Subtraction* of all unnecessary Expences; *Multiplication* (I would gladly have taught you that myself, but you thought it was time enough & wouldn't learn) he will soon make you a Mistress of it."

After a leisurely stay at Newport, Caty and Franklin resumed their journey to the Ward farm at Weekapaug, five miles southeast of Westerly. In a few days Caty was suddenly recalled to her Island home by the illness of one of her parents. Franklin wrote:

> I stood on the shore, and looked after you, till I could no longer distinguish you, even with my glass; then returned to your sister's, praying for your safe passage. Towards evening all agreed that you must certainly be arrived before that time, the weather having been so favourable; which made me more easy and cheerful, for I had been truly concerned for you. [March 4, 1755.]

On January 20 shortly after her return to the Island, Caty wrote her first letter to Franklin; unfortunately it is lost. He replied March 4, three days after his arrival at Philadelphia, with the letter quoted in part above. She wrote three letters more at intervals of about four weeks. Only the last, dated June 28, survives. Franklin was busy with supplying Braddock's army and failed to reply until September 11 in a letter which set the tone of their future correspondence.

> Philad[a] Sept. 11: 1755.
>
> Begone, Business, for an Hour, at least, & let me chat a little with my Katy.—
>
> I have now before me, my dear Girl, three of your Favours, viz. of March the 3[d] March the 30th. & May the 1st. The first I received [torn] before I set out on a long Journey, and the others while I was on that Journey, which held me near Six Weeks. Since my Return, I have been in such a perpetual Hurry of publick Affairs of various kinds, as render'd it impracticable for me to keep up my private Correspondencies, even those that afforded me the greatest Pleasure.
>
> You ask in your last, How I do, & what I am doing, and whether every body loves me yet, and why I make 'em do so? In the first Place, I [torn] Thanks to God, that I do not remember I was ever better. I still relish all the Pleasures of Life that a temperate Man can in reason desire, & thro' Favour I have them all in my Power. This happy Situation shall continue as long as God pleases, who knows what is best for his Creatures, and I hope will enable me to bear with Patience and dutiful Submission any Change he may think fit to make that is less agreable. As to the second Question, I must confess, (but don't you be jealous) that many more People love me now than ever did before: For since I saw you, I have been enabled to do some general Services to the Country, and to the Army, for which both have thank'd and prais'd me; and say they love me; they *say so*, as you us'd to do; and if I were to ask any Favours of them, would, perhaps, as readily refuse me: So that I find little real Advantage in being belov'd, but it pleases my Humour.
>
> Now it is near four Months since I have been favour'd with a single Line from you; but I will not be angry with you, because 'tis my fault. I ran in debt to you three or four Letters, and as I did not pay, you would not trust me any more, and you had some Reason: But believe me, I am honest, and tho' I should never make equal Returns, you shall see I'll keep fair Accounts.—Equal Returns I can never make, tho' I should write to you by every Post: For the Pleasure I receive from one of yours, is more than you can have from two of mine. The small News, the domestic Occurrences among our Friends, the natural Pictures you draw of Persons, the sensible Observations and Reflections you make, and the easy chatty Manner in which you express every thing, all contribute to heighten the Pleasure; and the more, as they remind me of those Hours and Miles that we talk'd away so agreably, even in a Winter Journey, a wrong Road, and a soaking Shower.
>
> I long to hear whether you have continu'd ever since in that Monastery [Block Island]; or have broke into the World again, doing pretty Mischief;

how the Lady Wards do, and how many of them are married, or about it; What is become of Mr. B. & Mr. L. and what the State of your Heart is at this Instant?—but that, perhaps I ought not to know; and therefore I will not conjure, as you sometimes say I do. If I could conjure, it should be to know what was that *oddest Question about me that ever was thought of*, which you tell me a Lady had just sent to ask you.

I commend your prudent Resolutions in the Article of granting Favours to Lovers: But if I were courting you, I could not heartily approve such Conduct. I should even be malicious enough to say you were too *knowing*, and tell you the old Story of the Girl & the Miller.

I enclose you the Songs you write for, and with them your Spanish Letter with a Translation. I honour that honest Spaniard for loving you: It show'd the Goodness of his Taste and Judgment. But you must forget him, and bless some worthy young Englishman.

You have spun a long Thread, 5022 Yards! It will reach almost from Block Island hither. I wish I had hold of one End of it, to pull you to me: But you would break it rather than come. The Cords of Love & Friendship are longer and stronger, and in Times past have drawn me farther; even back from England to Philadelphia. I guess that some of the same kind will one day draw you out of that Island.—

I was extreamly pleas'd with the Turff you sent me. The Irish People who have seen it, say, 'tis the right sort; but I cannot learn that we have anything like it here.—The Cheeses, particularly one of them, were excellent: All our Friends have tasted it, and all agree that it exceeds any English Cheese they ever tasted—Mrs. Franklin was very proud, that a young Lady sho[d] have so much Regard for her old Husband, as to send him such a Present. We talk of you every time it comes to Table; She is sure you are a sensible Girl, and a notable Housewife; and talks of bequeathing me to you as a Legacy; But I ought to wish you a better, and hope she will live these 100 Years; for we are grown old together, and if she has any faults, I am so us'd to 'em that I don't perceive 'em, as the Song says,

Some Faults we have all, and so may my Joan,
But then they're exceedingly small;
And now I'm us'd to 'em, they're just like my own,
I scarcely can see 'em at all,
My dear Friends,
I scarcely can see them at all.—

Indeed I begin to think she has none, as I think of you. And since she is willing I should love you as much as you are willing to be lov'd by me; let us join in wishing the old Lady a long Life and a happy

With her respectful Compliments to your good Mother & Sisters, present mine, tho' unknown, and believe me to be, D[r] Girl,

Your affectionate Friend
& humble Servant
B Franklin

Sally says, Papa, my Love to
Miss Katy.
If it was not quite unreasonable I should desire you to write to me every Post, whether you hear from me or no. As to your Spelling, don't let those laughing Girls put you out of Conceit with it. 'Tis the best in the world, for every Letter of it stands for something.—

During their sojourn at Boston, Caty and Franklin had become very good friends. She seems to have fallen into a habit of telling him, in confidence, about her many suitors and making personal remarks about them. Upon her return to the Island she began to think over what she had written and became apprehensive of having been indiscreet. Her first letter to survive, written June 28, 1755, reveals her uneasiness in a question as to what had become of her earlier letters. An unpublished fragment of an undated letter from Franklin sheds light on their confidential relationship and makes it clear that the cause of her apprehension was the fear that her correspondence might be seen by others. In the fragment Franklin assures her that her fears are groundless, that her letters will be seen by nobody but himself. When Franklin learned that he was being sent to London by the colony of Pennsylvania, he wrote Caty to say goodbye.

For the next five years there was no communication between them, so far as is now known. She may have attempted to tell him of her approaching marriage, for Franklin wrote Mrs. Franklin that a letter which she had forwarded had never been received. In 1758 she took Franklin's advice to "bless some worthy young Englishman" and became the wife of William Greene. Before Franklin's return in 1762 she had borne two of her six children.

When Caty heard of Franklin's being at Philadelphia, she wrote to congratulate him on his safe arrival and to invite him to visit Mr. Greene and her at Warwick should he be coming to New England.

Early in the spring of 1763 Franklin began another inspection tour of the post offices which took him from Virginia to Portsmouth, New Hampshire, a round trip of some sixteen hundred

miles. His daughter, Sally, joined him at New York, and together they proceeded towards New England. They travelled in a light carriage accompanied by a saddle horse which they rode in turn. Franklin was an indifferent horseman. He had had two bad falls which caused him much pain and from which he had not recovered a year later. About July 17 they arrived at Caty's in Warwick, Rhode Island, where they seem to have passed two days before proceeding to Providence en route to Boston and the Northeast. On the return journey in October they again spent a night at Mrs. Greene's and stopped with other friends along the road to Philadelphia where they arrived November 5. Near the middle of February, 1764, Franklin wrote to Caty to commiserate with her on the death of her mother and referred with pleasure to the kindnesses Sally and he had received on their journey "and particularly at your house."

For the eleven years Franklin was in England from 1764 to 1775, there has been found but a single fragment of a letter to Caty. It might be supposed that the correspondence had lapsed, as it apparently did, when Franklin was away between 1757 and 1762. But from the casual nature of the fragment, the fact that it deals with the commonplace topic of cheese-making, one is almost compelled to believe that Caty and Franklin were in more or less regular communication.

The loss of his letters is serious, but the loss of hers is a minor catastrophe; for there occurred in the year 1774 two events of importance in her life which she probably described to Franklin: the marriage of General Nathanael Greene to her niece, Catharine Littlefield, in the west parlor of her home, and her own baptism a few days later by the Reverend Ezra Stiles at the Court House in East Greenwich. Each occasion made a perfect subject for one of Caty's characteristic letters. Unfortunately we have only the official record made by Elder John Gorton, an invitation in Greene's own hand, and his letter to Sam Ward listing those who were to be present, from which to reconstruct the wedding. Of the baptism, Stiles's *Diary* gives a rather full account, but Caty's own reasons for being baptized at the ripe age of forty-three would have been most revealing.

The eighteen months between the arrival of Franklin at Philadelphia in May, 1775, and his departure for France in October, 1776, were among the busiest of his active life. He was immediately elected delegate to the Second Continental Congress and made a member of ten or more committees. Yet in this exciting period he had two good visits with his old friend, Catharine Greene; one at Caty's home in Warwick in October, 1775, and the last in Philadelphia in the late summer of 1776.

Late in May, 1775, he learned from John Adams that his sister, Jane Mecom, had fled from Boston and was at Warwick. "I suppose it must be at good Mr. and Mrs. Greene's, to whom present my affectionate respects," he wrote her. In October he went to Cambridge as a member of a committee to consult with Washington on the condition of the army. About the twenty-third he started for Warwick in a carriage which he had purchased for the purpose of taking his sister to Philadelphia. He found it possible to tarry at Caty's for four or five days and then set out for Philadelphia accompanied by Jane and Ray Greene, Caty's ten-year-old son. On the way they stopped at Governor William Franklin's in New Jersey as recorded in a letter from Jane to Mrs. Greene.

Near the end of July, 1776, Caty's husband, now Judge William Greene, was ordered by the General Assembly of Rhode Island to proceed to Philadelphia to receive the money due to the colony from the Congress for the troops taken into Continental service. Mrs. Greene left with her husband about July 22; they returned September 27, bringing with them in a chaise $120,000 of hard money. They were absent from Rhode Island sixty-eight days, but Judge Greene submitted the following account: "To my time in completing said Business being 46 days—exclusive of my taking ye smallpox." Caty's letter of October 1 to Franklin reported that the Greenes were ten days upon the road coming back so it may be assumed that the journey down to Philadelphia took about as long; therefore they were in Franklin's neighborhood for seven weeks, for three of which her husband lay sick of smallpox. From her letter one cannot say definitely that they were guests of Franklin's, yet it would seem probable, as he had visited the Greenes at Warwick upon three occasions. In any event she was in close proximity to Franklin for a longer period than at any other time in their twenty years of friendship. "Comeing from your house at first I hardly knew whether to be Glad or Sorry you was not at home as the Parting from those we love is Sorrow. . . ." This was their last meeting. Franklin sailed from Philadelphia

October 27 and arrived in France December 3, 1776.

News of the surrender of Burgoyne's entire army at Saratoga reached Franklin in Paris, December 4, 1777, when the American cause was at its lowest ebb. By the same vessel came two private letters to Franklin telling the great news, one from Elizabeth Partridge and one from Catharine Greene. It was the first that Caty had written since Franklin's departure a year before, and it brought him up to date on the family news; the most important item was that Mrs. Mecom had come out of Philadelphia a few days before Howe occupied the city and was now at Coventry, Rhode Island, with her granddaughter, "Jenny" (Flagg) Greene.

Four months later Franklin replied:

My dear old Friend, Paris February 28, 1778

Don't be offended at the Word *old;* I don't mean to call you an *old Woman;* it relates only to the Age of our Friendship; which on my part has always been a sincerely affectionate one, and I flatter myself the same on yours. . . .

We are all glad to hear of Ray, for we all love him.—[He was at school in Newbury, Mass.] I have been often much concern'd for my Friends at Warwick, hearing that the Enemy was so near them. I hope your Troubles will not be of much longer Duration: For tho' the Wickedness of the English Court, & its Malice against us is great as ever, its Horns are shortened; its Strength diminishes daily; and we have form'd an Alliance here, & shall form others, that will help to keep the Bull quiet, and make him orderly.—I chat, you see, as usual, anyhow, with you, who are kind enough never to criticise Improprieties in my Comparisons or anything else.— . . . I pity my poor old Sister, to be so harass'd & driven about by the Enemy. For I feel a little myself the Inconvenience of being driven about by my Friends.—I live here in great Respect, and dine every day with Great Folks; but I still long for home & for Repose; and should be happy to eat Indian Puding in your Company & under your hospitable Roof.

To this last Jane replied on August 15, and referred to the battle then preparing on Rhode Island:

I hope they [the American and French forces] will have there desiered Suckses for the Sake of the whol comunitie & a litle for my own for I have lived in constant Jeperdie Since the Spring when my children removed from Coventry to this place where we are much Exposed & have been under constant Aprehensions, I have been Part of the time at the Governers [William Greene] but it was full as bad there for they offered a reward for Takeing him, you will Acknolidg this is Rather wors[e] than being harrised about by wons Friends, yet I doubt not but that is Trobleson to you who are So desierous of Retierment, I fear you will never be Suffered to Injoy it.

Franklin's achievements at the Court of France aroused the envy of his enemies who did not hesitate to use any means to defame him. His success with the ladies was a prime target; even Jane did not hesitate to tease him about it:

I have not yet recd a line from you since that by Mr Simeon Dean, but bless God I now & then hear of yr helth & Glorious atchievments in the political way, as well as in the favour of the Ladys ("Since you have rubd off the Mechanic Rust and commenced compleat courtier) who Jonathan Williams writes me clame from you the Tribute of an Embrace & it Seemes you do not complane of the Tax as a very grat penance. [June 23, 1779.]

Playfully Caty touched on the same topic:

Will you believe I grow Very Jealous of you I fear the french Ladies have taken you intirely from us for we dont have a Single line from you this long Very long time the last letter I had was Recommending the two French Gentlemen. [October 7, 1781.]

The British evacuated Newport, October 25, 1779, after occupying it for thirty-five months; Count Rochambeau arrived with 6,500 French troops July 10, 1780. At the end of the year Caty wrote Franklin:

I know it will give you Pleasure to hear we are Well, And have not fell in to the hands of the Britons as we have been fearfull we Should living So near the Shore and Spoues So Warm a Whigg—but them fears are at an End Since We are So Strongly guarded by our New Allies. Who we are much Pleasd With there Politeness & Strong Atachment to America is Very indearing—Count Rochambo We Particularly Regard the more as he Says he loves Us and he is Clever [amiable] to every body—Admiral Turney [De Ternay] Who Died in Newport a few days Past We have Reason to Lament As he was exceedingly kind to our Son Ray who has been on Board his Ship more than 2 months to learn French. . . .

The uncertainty of a letters Reaching you makes me Seldom write but be Assurd I love you as Dearly as ever And hope that Providence will favor us with yr Safe Return to New England Where we Will Welcom you with our Hearts & our Hands and Sugar Plumbs too. . . . [Dec. 28, 1780.]

Since late in 1777 Jane Mecom had been living with her granddaughter, Jane (Flagg) Greene,

who alternated between Potowomut and Coventry, remote from close neighbors. Jane had complained to her brother July 27, 1779: "I feel the want of suitable conversation I have but little hear." Caty did her best to amuse her friend: "When we have the Emtyest house We Send for her to Stay with us She Cant Bear Company as She used to do We love her and She loves us," she wrote Frankin September 19, 1779.

At last Franklin came home, landing at Philadelphia, September 14, 1785. Within the week he wrote to Governor and Mrs. Greene:

Philad^{a} Sept. 20. 1785.

I seize this first Opportunity of acquainting my dear Friends that I have once more the great Happiness of being at home in my own Country & with my Family, because I know it will give you Pleasure. I shall be glad to hear of your Welfare also, and beg you to favour me with a Line, and let me know particularly how my young Friend Ray does.— I enjoy, Thanks to God, as much good Health as can reasonably be expected at my Time of Life: and am ever, with sincere Esteem, my dear Friends

Yours most affectionately

B.Franklin

Gov^{r} Greene & Lady

Caty replied October 8: "Welcom Exceedingly Welcom my Very Dear Friend to america, the Pleasing news gave me a Spring I Cant Describe, happy your Friends and your family Particularly. May you injoy health and every Blessing to a Very great age." Jane wrote him: "our Friend Catharine Greene Expresd such Lively Joy at the News of yr Arival that her children tould Her it had thron her in to Histericks but She Says She is not Subject to that Disorder."

Much as he wished to pay a visit to New England and much as his "two old women" prayed that they might see him, Franklin's health prevented his stirring from Philadelphia. Walking was painful and riding in a carriage excruciating, so he was usually carried in his sedan chair. A fall on the stone steps of his garden late in 1787 greatly aggravated the pain caused by the stone in his bladder, and by the end of 1788 he was obliged to resort to opium for relief; this destroyed his appetite and he became emaciated. For the last year of his life he never left his bedroom. Neither Jane nor Catharine ever saw him after October, 1776.

December 10, 1787, Catharine wrote in what seems to be her last letter to Franklin that she was sending him another bag of sweet corn in care of Richard Bache in New York:

I left a few lines with yr good Sister when in Boston to forward to you of how do you do &c. we had each of us a feast talking of you But Some aloy when you give no incoragement of Comeing to New England how glad it would make many many hearts to See you how is yr health We hear from you by many People that you are Chearfull and Pleasant as may, but I long to hear from yr own hand that you are Comfortable but when I reflect the many Corespondents you have the multiplicity of Busines, I dare not Please my Self with the hope.

Franklin replied on March 2, 1789:

Philad^{a}

Dear Friend,

Having now done with Public Affairs which have heretofore taken up so much of my Time, I shall endeavour to enjoy, during the small Remainder of Life that is left to me, some of the Pleasure of Conversing with my old Friends by Writing, since their Distance prevents my Hope of seeing them again.

I received one of the Bags of Sweet Corn you were so good as to send me a long time since, but the other never came to hand; even the Letter mentioning it, tho' dated Dec. 10, 1787, has been above a Year on its way, for I receiv'd it but about two Weeks since from Baltimore in Maryland. The Corn I did receive was excellent, and gave me great Pleasure. Accept my hearty Thanks.—

I am, as you suppose in the above mention'd old Letter, much pleas'd to hear that my young Friend Ray is *"smart in the Farming Way"* and makes such substantial Fences. I think Agriculture the most honourable of all Employments, as being the most Independent. The Farmer has no need of popular Favour, nor the Favour of the Great. The Success of his Crops depending only on the Blessing of God upon his honest Industry. I congratulate your good Spouse that he as well as my self is now free from public Cares, and that he can bend his whole Attention to his Farming, which will afford him both Profit and Pleasure; a Business which no body knows better how to manage with Advantage. I am too old to follow Printing again my self, but loving the Business, I have brought up my Grandson Benjamin to it, and have built and furnish'd a Printing-house for him, which he now manages under my Eye. I have great Pleasure in the rest of my Grand children who are now in Number eight, and all promising, the youngest only 6 Months old, but

shows signs of great Good nature. My Friends here are numerous, and I enjoy as much of their Conversation as I can reasonably wish; and I have as much Health & Chearfulness as can well be expected at my Age, now 83. Hitherto this long Life has been tolerably happy,* so that if I were allow'd to live it over again, I should make no Objection, only wishing for Leave to do, what Authors do in a second Edition of their Works, correct some of my Errata. Among the Felicities of my Life I reckon your Friendship, which I shall remember with Pleasure as long as that Life lasts, being ever, my dear Friend,

Yours most affectionately

B Franklin

* (God grant it may so continue to the End)

WHY UNION EDUCATION?

Aims, History, and Philosophy of the Educational Work of the International Ladies' Garment Workers' Union

MARK STARR

Educational Director, International Ladies' Garment Workers' Union

(*Read November 21, 1947*)

WHEN Benjamin Franklin, in 1727, organized the first Junto in Philadelphia, which was perhaps the earliest effort in adult education in this country, he put among the weekly ritual questions: "Have you lately observed any encroachment on the just liberties of the people?" Another query to provoke reports and discussion was: "Have you lately observed any defect in the laws of your country of which it would be proper to move the legislature for an amendment?" Or "Do you know of any beneficial law that is wanting?" [1]

So you can see that the founder of your own Society not only was wise enough to reject the deadly heresy that education was confined to formal work in school and college but he also saw that education must be a continuing life-long process and further that it must include education for social action. Workers' education, as carried on by all unions and by the International Ladies' Garment Workers' Union in particular, shares that motivation. Its emphasis is upon the study of group problems to result in group action for their solution.

The urgent nature of the early problems of the needle trades can easily be seen in the recollections of Pauline Newman, now directing educational activities for the Union Health Center. She recalls the work conditions when the ILGWU began in 1900:

> I well remember the factories being located in basements and backyards; no proper ventilation, hardly any illumination except gas jets burning from morning till night, no toilet facilities except in the yards. Filth and dirt were a part of the industry. I remember working from 70 to 80 hours a week, seven days a week; the earnings ranged from $1 to $3 for youngsters to $5 and $6 for operators, and skilled mechanics did not earn more than $25 a week. The factories were very dangerous indeed—they were practically fire traps. We endangered our lives every day we spent there.

Such conditions led to the establishment in the industry jointly by the employers and the union of the Joint Board of Sanitary Control and later to one of the finest health institutions in the country—the Union Health Center, which in 1946 rendered some 207,880 services—more will be given in 1947. The cost of the Center has been borne by the Union. In recent years employers have also contributed to far reaching health and welfare plans administered by the union. And union health centers have been established in Philadelphia, St. Louis, Fall River, Wilkes-Barre, Harrisburg, and other cities.

PURSUIT OF HAPPINESS

"All men are endowed by their Creator with certain inalienable rights . . . life, liberty and the pursuit of happiness," asserts the Declaration of Independence. And fighting for the "life" and "liberty" of its members and of the working class the world over is the job of the International Ladies' Garment Workers' Union. The goal of its Educational Department, established in 1917, is "the pursuit of happiness." [2]

By providing job security, decreasing the length of the work week, and raising wages, the union has given workers more leisure as distinct from the enforced idleness of unemployment. The Educational Department, founded to provide a fuller life for all ILGWU members and their families, has established facilities for the enjoyment of this leisure time. Through an extensive program of education and recreation, the Educational Department builds happier men and women, stauncher and stronger trade unionists. Its classes in economics, history, literature, and philosophy help members to understand the world and its problems. Through current events forums, debates, radio programs, and printed material, workers are prepared for full-time citizenship and

[1] Sparks, Jared, ed., *The works of Benjamin Franklin* II: 11, Boston, 1836.

[2] The material which follows is based on "And the Pursuit of Happiness," a pictorial history of the ILGWU Educational Department, New York City, 1947.

active participation in public affairs. Classes in labor problems and trade unionism train members for effective trade union careers. Opportunities in drama, music, the arts, and sports provide for cultural development and self-expression after the day's work.

The ILGWU Educational Department is a dynamic force in the life of its members because, in true American fashion, the courses and programs of the Department are shaped by the desires of the members of the union. It is proud of the fact that its pioneering work has influenced the educational programs of other unions as well as certain phases of general adult education. Today, as at its official inception in 1917, however, the Department's major concern is the enrichment of the lives of the workers. It shares Ruskin's belief: "There is no wealth but life—life with all its opportunities for joy, admiration and knowledge. That nation is the richest which has the greatest number of healthy, happy human beings."

Because it realizes that the garment workers are a part of the community and that no one group can be happy and secure by itself, the ILGWU, throughout the years has worked closely with other community-minded groups in behalf of all citizens. The union, for example, has added its strength to charity drives. The National Foundation for Infantile Paralysis, the American Red Cross, the Federal Council of the Churches of Christ in America, the Federation of Jewish Philanthropic Societies, individual hospitals, refugee agencies, and youth organizations receive the ILGWU's contributions regularly. Particularly has it joined with other groups to win reforms, for championing all legislation that will benefit the majority of Americans is traditional with the ILGWU. The union has been in the forefront of the fights for workmen's compensation, unemployment insurance, minimum wage laws, a permanent Fair Employment Practices Commission, the Murray-Wagner-Dingell bill, and full employment. At the same time, the union supports political campaigns designed to provide citizens with better government. At home and aboard, it cooperates with all labor groups and has contributed over one million dollars to striking AFL and CIO unions.

THE JOB OF RECONSTRUCTION

The union itself is alert to current changes in the social climate. For instance, on August 14, 1945, while 320,000 garment workers joined with their fellow-citizens throughout the nation in celebrating the victorious ending of World War II, the International Ladies' Garment Workers' Union began the equally welcome task of reconversion. The union carried forward its services to the community at large and made plans for the operation of FM radio stations designed to provide education and worth-while entertainment for ILGWU members and the general public in six key cities. Still another community project undertaken by the ILGWU is a veterans on-the-job training program to relieve unemployment among returning servicemen. The union has enlisted the cooperation of the Veterans Administration and a number of employers. As a result hundreds of veterans were trained for jobs in Fall River, Mass., and in other garment-industry centers.

The ILGWU Educational Department at the end of World War II also made plans for an entensive program of postwar education and recreation immediately after the end of hostilities. Special veteran-civilian discussion groups on postwar employment, housing, race problems, and peace were instituted in many cities and gave veterans and civilians a chance to share and exchange experiences. Simultaneously, the prewar schedule of educational, cultural and recreational activities has been reinstituted. Literature, economics, and history classes, lectures on social legislation and the international situation, choruses, orchestras, art classes, dramatics and dance groups flourish. Special courses on the wartime growth of trade unionism have been added to the regular classes provided and the Officers' Qualification Courses, which prepare ILGWU members for union office. Local union classes for new members, long a part of the curriculum, have become mandatory in such cities as Knoxville, St. Louis, Minneapolis, and Dallas, to insure an alert and intelligent postwar union membership. Returning veterans have revitalized old ball teams and organized new ones. The girls have replaced bandage rolling and canteen work with book clubs, lectures on international relations, and personality courses.

OUT OF THE DARK PAST

The current educational program is the outgrowth of a modest but steadfast policy of workers' education begun by the ILGWU in the early 1900's. That program, fashioned from the

dreams and visions of the pioneers who built the union, who dared to plan for the day when workers would have happiness and leisure as well as industrial rights, at first provided members with only a few simple activities and study courses. Then as now, however, the educational program was dedicated to one objective—the "pursuit of happiness" by garment workers and other workers everywhere. And in the early days, the workers were in desperate revolt against conditions which made happiness impossible. Sweatshops in the first years of our century were everywhere, perched in lofts over stables; concealed in dark, rat-infested cellars; stuffed into the bedrooms and dark hallways of slum dwellings. There were no washrooms, no drinking water; gas jets burned day and night. On July and August "dog days" the workers, carrying their work with them, sought fresh air on roofs and fire escapes. Older workers toiled from sixty to eighty hours, as Pauline Newman already reported earlier in this paper, for a weekly wage of $5.00 to $6.00. Young men and women, classified as "learners," worked for $1.50 to $3.00. The work week included Saturday and part-time on Sundays. In the busy season men frequently worked all night and put in a full day on Sunday. Pitted against each other, workers paid for the use of sewing machines, thread and needles, and the very chairs on which they sat.

It was to remove such conditions that on June 3, 1900, eleven delegates representing seven unions with a total membership of 2,000 workers in New York, Philadelphia, Baltimore, and Newark met in New York City to discuss the formation of a garment worker's union which would combat the evils of the garment industry and the sweatshops which sapped the strength of those who worked in them. Less than three weeks later, on June 23, the International Ladies' Garment Workers' Union was chartered by the American Federation of Labor.

Many of the ILGWU's members were Americans by choice. They had left homes in Russia, Poland, Italy, and Hungary because of the promise held out by the New World. Many had experienced the pogroms and massacres of Europe. Others had fled political persecution and still others had been teachers and leaders in the union movement abroad. Life in America proved to be something less than the "life, liberty, and pursuit of happiness" they sought. They became garment workers because they had to make a living, but their dreams of a New World and their thirst for knowledge persisted. One local of Italian-speaking workers designated itself Local 89 in commemoration of the French Revolution of 1789 and chose as its motto, *Pane e Rose* (Bread and Roses). This motto was, in essence, the motto of the entire ILGWU—for to those who made up the union the right to earn a living and the right to liberty and happiness were inseparably bound together.

The early garment workers looked upon education as the road to a fuller life. Many of them brought their devotion to socialist educational clubs over from Europe. Now, they naturally transferred the spirit of these clubs, which featured informal debates, forums, and public meetings on social, political, and economic questions, to their union halls. Two great strikes, the "Uprising of the 20,000" in 1909, and a year later, the "Great Revolt," in which 50,000 cloakmakers participated, paved the way for the introduction of formal education. Now, the union was sufficiently entrenched to begin fighting for the "intellectual and spiritual" emancipation of its members. For the first time in the history of the industry, a collective agreement was signed by the union and the employers' associations. The Protocol of Peace, formulated by a group of industrial statesmen headed by Louis D. Brandeis, provided for collective bargaining, the six-day week, ten legal holidays, and weekly payments in cash instead of promises and bad checks. It did away with many abuses—no longer did workers have to pay for their machines and thread; homework and inside subcontracting were abolished. It established a Joint Board of Sanitary Control to supervise sanitary conditions in the shops.

The tragedy of the Triangle Fire, which in 1911 wiped out the lives of 146 young girls, strengthened the union's determination to educate its members so that employers could no longer take advantage of their ignorance by making them risk their lives by working in sweatshops where factory doors were locked to shut out union organizers. Many social workers and far-seeing politicians assisted the union to secure laws for safety in the factory because they had been shocked by the holocaust.

NOT BY BREAD ALONE

The union grew in strength and attracted new members. And in 1913 Local 25 (New York Waistmakers) instituted courses in trade-union instruction as well as classes in English for immigrant members and, in a very real sense, launched the preliminaries of the ILGWU's Educational Department, which thirty years later was to have a budget of $200,000 and a staff of over thirty educational directors. A year later delegates to the ILGWU's convention urged the union to "dwell particularly on the preparatory work of education." President Benjamin Schlesinger arranged with the Rand School of Social Sciences for courses in the history and practices of trade unionism.

The 1916 convention officially established the ILGWU Educational Department and appropriated $5,000 for its work. Three years earlier the ILGWU had established the Union Health Center which today provides therapeutic and diagnostic service for 50,000 ILGWU members and their families each year. In 1916, too, the union turned its attention to the recreational needs of its membership. An ILGWU local, pioneering once more, established Unity House, the union's first summer resort for workers. Originally situated in the Catskills, Unity House, which set a pattern for workers' recreation, was later moved to the Pocono Mountains in Pennsylvania and taken over by the national office of the ILGWU. By 1917 the union was entrenched and the pattern to which it would adhere was established. From its very first days, it had taken an active interest not only in the economic problems but the health, educational, and recreational life of its members. Today this interest embraces the happiness and well-being of working people throughout the world.

During World War I, the garment industry shared in the general war prosperity and the ILGWU gained many new members as well as wage increases and shorter hours. The Educational Department, sharing the feeling of "aggressive hopefulness" that pervaded the union, in 1917 launched a full-time program. In New York City it opened a Workers University at Washington Irving High School. Public school evening classes called Unity Centers were sponsored by the Department. The teaching ranged from English and naturalization classes to labor problems, philosophy, and world literature. Similar programs were sponsored by the union in Philadelphia, Boston, Chicago, Cleveland, Worcester, Cincinnati, Montreal, and Newark.

WIDER HORIZONS

The membership responded so enthusiastically that a year later the AFL surveyed the educational program and urged other American labor unions to follow the trail blazed by the ILGWU and to establish their own educational departments. By 1924 the ILGWU had helped hundreds of garment workers to obtain their naturalization and citizenship papers. Then, because immigration had been drastically reduced as a result of the 1921 and 1924 immigration laws, the number of potential citizenship students decreased. Once again the Educational Department passed a milestone in its development and prepared to meet the changing needs of the members. To accommodate the change in its enrollment, the ILGWU recast the program of its Educational Department. While English classes were continued for those garment workers who required such instruction, the Workers University was enlarged and the class and lecture program was widened to include forums, physical training, literature, music appreciation, and health. Many of the accomplishments of the early 1920's had been made in the face of a severe postwar depression. When conditions began to improve, the union confidently looked forward to a period of growth and consolidation.

But in 1926 dissension broke out in the ILGWU. Its strength was drained from within when a group of Communists gained temporary control of the New York Joint Board and launched a disastrous and lengthy strike. When the strike ended, the treasury was empty and the union was seriously in debt. Before it had a chance to recover its strength and resources, there came an equally severe blow from the outside. The stock market crashed in 1929 and ushered in the greatest unemployment in the history of the nation. The garment industry weathered the first two years of depression, but by 1931 the industry was at a virtual standstill. Thousands of union members lost their jobs and it was difficult—if not impossible—to collect dues. Despite the unemployment, the ILGWU was

forced to levy per capita taxes on its locals in order to reduce its debt and make a slow start toward rehabilitating the union treasury.

ARBITRATION ESTABLISHED

One of the few causes for optimism came when, as a result of a 1930 dressmakers' strike, an arbitration system with a full-time impartial chairman was established. This system remains today a model for the industry. The 1930 strike was quickly settled by the intervention of Governor Franklin D. Roosevelt and Lieutenant Governor Herbert H. Lehman. The union did not fully recover, however, until the advent of the New Deal in 1933 when federal legislation was passed stabilizing business and industry. The passage of the NRA (National Recovery Act) which made it illegal for employers to oppose labor organizations and outlawed the company union gave trade unionism the government support it needed to assume its rightful place in American industry. Seizing its opportunity, the ILGWU launched a vigorous and successful organizing drive covering sixty cities in the United States and additional cities in Canada. Years of experience went into a campaign planned in advance to the last detail, and the ILGWU doubled, then tripled, and finally quadrupled in size until in 1937 its membership totaled 225,000. These new members, promptly christened "NRA Babies," greatly changed the character of the membership. They were for the most part young American-born girls representing many different national and racial origins but reared in the American scene and in smaller cities previously untouched by union organization.

The NRA was declared unconstitutional but the National Labor Relations Act which succeeded it enforced the workers' right to organize, and the ILGWU membership continued to grow.

TRADE UNION EDUCATION

On the organizing picket lines of the Blue Eagle period new labor songs and plays were born. And, a new spirit pervaded the union. The Educational Department became more important than before. In the words of President David Dubinsky, the Department now had to "weld and keep together the masses who until now were strangers to trade-union organization." It was important to give the "NRA Babies" a knowledge of the history, methods and ideals of trade unionism in general and of their own union in particular. As members and potential officers, they needed to be fully informed about their rights and duties. They needed to know how to run meetings, settle grievances, negotiate contracts, and organize their fellow-workers. In the economics classes the worker learns that the huge plants found in most important American industries are rare in the garment trades. The average garment shop employs approximately thirty-five to forty workers. Another lesson makes it clear that the ILGWU has jurisdiction over the manufacture of dresses, coats and suits, underwear and nightwear, corsets and brassieres, infants' and children's wear, artificial flowers, plastic or metal novelties used for dress ornaments, buttons, embroidery, neckwear, and rainwear. Garment workers also learn that under agreements with the employer or his association and the union, the worker is assured a definite price for his or her labor and a fair share of the available work. The union's price committees and the employer's representative compute rates for each operation on every new design so there can be no unfair competition between workers in various shops and districts. Before the war, in the New York dress industry alone, piece rates were computed on more than 100,000 different styles a year. Little imagination is needed to recognize the skills and alertness necessary for union officers and workers to settle prices on such complicated and numerous operations.

Because they were young men and women brought up with the American love of sports and recreation, the new members of the New Deal decade 1933–43 also wanted recreation and social programs to be made available to them. Varied cultural and recreational activities including lectures, trips to places of interest, sports, music, art, and drama were made accessible in ILGWU centers throughout the country. Study courses in economics, labor problems, social history, and history of the labor movement as well as English, parliamentary law, and public speaking reached record-breaking levels of attendance in ILGWU centers throughout the country. Because the strength of the union depended on a well-informed membership, trade union training was stressed above all else. This special training program included classes for new members; courses for shop chairmen, local executive board members, and potential officers; instruction in the structure and functioning of the ILGWU; the techniques used

in organizing, negotiating, conducting union meetings, keeping union records, striking, and presenting employee demands. These courses also stressed the economics of the garment industry—a disorganized and small-scale industry stabilized to a great extent by the union in spite of the numbers of designs, the rapidity of style changes and its seasonal fluctuations.

The ILGWU's *Handbook of Trade Union Methods* was issued in 1937 for the benefit of these members. It became a best seller. Two editions were used by other unions and teachers of labor problems. A special edition was also issued for the United Automobile Workers of America, CIO. Other pamphlets, lesson syllabi and books were published for use in ILGWU and other union study groups.

The idea of on-the-job training, never before attempted by a union, worked. The "NRA Babies" became staunch and steadfast trade unionists. And, the convention of 1937, impressed with the keener understanding shown by officers and members as a result of union education, amended the ILGWU Constitution, making it compulsory for all new candidates for full-time union posts to complete the Officers' Qualification Courses.

From the first, workers acquired new respect for themselves and their jobs. In the words of the ILGWU anthem, born in the New Deal decade:

One battle is won, but the fight's just begun,
And the Union flag's unfurled—
United we're strong, let us march toward the dawn
Of a brave new workers' world.

RELATIONS WITH HIGHER EDUCATION

The ILGWU had successfully taught trade unionism to newcomers in the labor field. In an attempt to reach the general public, in 1937 and 1938 it sponsored a college essay contest on "The Legal Responsibilities of Trade Unions." Open to junior and senior students in all accredited American colleges, it offered over $1,500 in prize money. The union has always maintained cordial relationships with schools and colleges, supplies information and speakers to, and receives visits from, educational, religious, and civic groups.

Scholarships were provided by the union for ILGWU members who wished to attend workers' summer schools at the Hudson Shore Labor School and the University of Wisconsin. (See appendix p. 202 for details.)

In 1941 the ILGWU Educational Department cooperated with Harvard University in establishing fellowships for union members. The holder of a fellowship is entitled to eight months of intensive study at Harvard. Women were made eligible for these fellowships in 1945 at the request of the ILGWU, and two Negro women were among the recipients.

ON STAGE AND IN PRINT

During this period, the union's cultural and recreational activities reached full maturity. The production of the musical revue, *Pins and Needles*, aroused nationwide attention. *Pins and Needles*, first presented at Labor Stage, a small theater acquired by the ILGWU in 1936, played to capacity audiences for two years on Broadway, made a highly successful cross-country tour, and served as a good-will ambassador for the ILGWU and trade unions in general. Labor Stage also provided studios for ILGWU dance groups, orchestra rehearsals, and choral meetings. A second dramatic offering, *I Hear America Singing*, was presented at the opening session of the 1940 convention and played a repeat performance to a capacity audience at Madison Square Garden. In the field of fine arts enthusiastic reviews acclaimed the art exhibit sponsored by ILGWU art students in New York, New Jersey, and Canada. The public concerts given by the ILGWU chorus and mandolin orchestra were received with similar enthusiasm.

At the same time the ILGWU recreational program, which had developed rapidly since 1933, was expanded to include almost all sports. Locals sponsored baseball, basketball and soccer teams; individual garment workers took up tennis, handball, golf, swimming, archery, horseback riding, wrestling, and attended ILGWU-sponsored bowling and regular gym classes.

Justice, the ILGWU newspaper, is published bi-monthly and reaches 350,000 member subscribers. *Gerechtigkeit* is published for members who read Yiddish, *Giustizia* for those who read Italian, and *Justicia* for those who read Spanish. Montreal members run a journal in French.

UNITY HOUSE

The union's summer home, Unity House, played an integral role in this cultural and recreational program. Then as now, this million-dollar vacation center, located on 850 acres in the Pocono Mountains and operated on a non-profit basis, provides for the best in vacation facilities, entertainment, and recreation for unionists and their guests, at rates which all members are able to afford. Each summer outstanding lecturers, concert artists, dancers, and sport champions make use of the open-air theater, the lake, tennis courts, and lecture halls for the entertainment of ILGWU members and their guests.

WARTIME CHANGES

When fascism first began to gain a foothold in Europe, the ILGWU aligned itself with anti-fascist and anti-Nazi forces. And, when the United States went to war on December 7, 1941, the union threw itself whole-heartedly into the war effort. On the job, it devoted its skills to the production of uniforms for nurses, Red Cross workers and women members of the armed forces, as well as to the making of cartridge belts, parachutes, and other war necessities. In their spare time members sold millions of dollars' worth of War Bonds, made garments for bomb victims, donated blood, raised hundreds of thousands of dollars for relief work, rest homes, and schools. The union provided funds for the European and Chinese underground. The Educational Department likewise shifted the focus of its program from peace to war and organized numerous activities through which ILGWU members participated in the war effort. Special classes in first aid, home nursing, nutrition, consumer problems, and knitting were added to the curriculum. A Women's Service Brigade, to which workers were automatically eligible upon the successful completion of the first aid, home nursing, and nutrition courses, was organized in New York City to aid relief campaigns and other civilian wartime drives. St. Louis, Boston, and Kansas City locals ran canteens for servicemen. In New York, Labor Stage was converted into a week-end canteen, which before its lease expired in 1943 entertained 10,800 members of the armed forces of the United Nations. The Department bought and distributed five thousand dollars' worth of books to equip the libraries of the Liberty Ships "Meyer London," "Morris Hillquit," "Benjamin Schlesinger" and "Morris Sigman," built by ILGWU bond subscriptions.

It was difficult to hold formal classes in wartime. Many activity projects run were difficult to measure in statistical reports and class attendance fell. Many members of the union joined the armed forces; attendances were less regular owing to overtime, transportation difficulties, alternating shifts, and the distractions resulting from war strain. Nevertheless, in 1942, 17,000 garment workers were enrolled in ILGWU classes. Officers' institutes continued and featured such speakers as Henry J. Kaiser, Elmo Roper, Eric A. Johnston, and William H. Davis.

The department also expanded its lecture service, and at bi-weekly intervals packets of pamphlets and other materials were supplied to all local educational committees. A long-established book purchasing service which makes it possible for members to obtain books at minimum cost was enlarged. A movie service called "Movies for Morale" supplied wartime films to locals free of charge. Even the cultural activities sponsored by the Educational Department—restricted but not completely curtailed during the war period—were diverted toward the war effort. The concerts given by the ILGWU Symphony Orchestra in 1942 and 1943 helped to sell hundreds of thousands of dollar's worth of War Bonds. Traditionally, the ILGWU has always played a prominent role in community life and community affairs. Its wartime cooperation with the community was an intensification of its normal activity. It also naturally cooperates with labor groups because it realizes that the security of its members depends on the security of all labor, and that no group can be well-housed, well-fed, and prosperous for long unless all workers are. For this reason, the ILGWU works with other labor unions, both in this country and abroad. It endorses progressive political programs, encourages independent political action, and lends its support to organizations whose goal is civic and political education.

RELATIONS WITH MANAGEMENT

The ILGWU has blazed a trail of enlightened labor-management relations. Since 1933 there have been no large-scale strikes in the garment industry. Differences are settled by men seated at conference tables; contracts are re-negotiated while work goes on in the shops.

The union has initiated and participated in many projects to increase the prosperity and stability of the garment industry. Among these are the New York Dress Institute which promotes general dress and fashion consciousness and the National Coat and Suit Recovery Board, which encourages fair labor standards. Its Management Engineering Department has established standard techniques for determining fair piece rates and has saved thousands of dollars for employers and employees alike by streamlining manufacturing operations and methods.

The union's Research Department, designed primarily to assist the union, is also ready to extend a helping hand to the industry. The only center of its kind, the Department watches conditions throughout the industry and can, at a moment's notice, provide information on the number of dress styles on the market in any price range, the number of women's coats that are sold by a department store during a given year, market reactions, and general economic conditions.

Welfare work providing medical care, hospitalization, old age and sickness and health benefits now cover the majority of the members and these are coordinated by the union's Health and Welfare Benefits Department.

EDUCATION FOR TOMORROW

So much for the past. What about the future? The ILGWU believes that today the pursuit of happiness depends more than ever on wiping out social illiteracy and attaining civic and political maturity. Education for freedom and education for social planning must be provided so that the peoples of the world may understand each other and together control the forces of the atom. Citizens of this nation must be prepared for active participation in civic life so that the scourge of fascism may never again be revived. Simultaneously, there must be planning for an era in which technological progress will insure additional leisure, and for the extension of facilities for the wholesome enjoyment of this new opportunity for fuller living.

All of these goals may be reached through education—an education program whose mission is not only the preparation of young men and women for business and industry, but also the preparation and training for living in its broadest social implications.

The pursuit of happiness has no final goal. Each day brings the hope of further advance toward the achievement of a better world—a world of decent jobs for all; a world in which children will have fair and ample educational opportunity; in which wars and their terrors will have disappeared. The pursuit will continue so long as life lasts. And the International Ladies' Garment Workers' Union will continue through group solidarity and action in industry, politics, and education to be in the vanguard of the march forward.

APPENDIX A

ILGWU "FIRSTS"

- Union Health Centers
- Educational Department
- Vacation resort for workers
- Management Engineering Department
- Research Department
- In-service training for union officials
- Publication of union handbooks
- Published financial statements covering both general and local offices

APPENDIX B

WAR RECORD

ILGWU members—

- Gave $6,500,000 for war relief
- Sold $170,000,000 worth of War Bonds
- Made 50,000 garments in their leisure time for liberated Russians
- Established rest homes for women workers in Britain; maintained a Merchant Navy Club in the heart of London
- Provided nurseries and houses for bombed-out English children
- Contributed $10,000 worth of medical supplies and $75,000 to Russian relief
- Contributed $100,000 to the Histadrut for building "International House," a training school for youth in Haifa, Palestine
- Contributed $250,000 for the building and support of a vocational training school for children in Rome
- Gave $150,000 to China to build a vocational school for war orphans on the outskirts of Chungking
- Furnished $100,000 worth of clothing for Polish women victims of the Nazis

APPENDIX C

SUMMARY 1935–1946

Classes and Groups

	Locals and Jt. Bds. Participating	Study Classes	Music	Dramatics	Athletics	Gym	Dancing	Total Groups	Total Participants
1935–36	59	293	55	33	51	54	21	507	15,000
1936–37	76	317	44	37	51	71	33	553	20,000
1937–38	75	309	70	37	37	84	37	620	22,050
1938–39	74	268	117	74	59	100	59	673	16,464
1939–40	76	261	120	80	50	142	50	703	17,000
1940–41	79	290	103	42	41	172	56	704	16,800
1941–42	87	494	102	30	116	168	58	968	19,887
1942–43	82	418	56	19	126	52	34	705	17,235
1943–44	72	265	45	10	107	39	23	489	11,878
1944–45	56	112	28	5	49	24	15	233	6,840
1945–46	72	152	30	7	60	32	25	306	8,962

STUDY CLASSES

Trade Union History
Trade Union Techniques
Arts and Crafts
Health and Nutrition
Esperanto
Parliamentary Law
Public Speaking
Psychology
Consumer Classes
Labor Problems
Music Appreciation
Current Events
Economics
English

INSTITUTES AND SCHOLARSHIPS

	ILGWU Institutes		Scholarships at Labor Schools
Year	Number	Students	
1935–36	8	280	11
1936–37	5	164	23
1937–38	3	140	17
1938–39	4	208	12
1939–40	4	205	3
1940–41	4	230	5
1941–42	5	264	4
1942–43	3	185	8
1943–44	3	220	6
1944–45	7	465	5
1945–46	10	476	11

SOME BASIC ISSUES RAISED BY TRADE UNIONS

SUMNER H. SLICHTER

Lamont University Professor, Harvard University

(*Read November 21, 1947*)

I

TRADE union membership in the United States today includes nearly half of the non-supervisory and non-technical employees in private industry. Two-thirds of the workers employed in manufacturing, four-fifths in transportation, four-fifths in mining, and four-fifths in construction are organized. Trade unions vary greatly in strength, but the most strategically situated are undoubtedly the most powerful economic organizations which the country has ever seen. So widespread and powerful are unions that they are bringing into existence a new type of economy—the collective bargaining economy of the twentieth century is succeeding the free market economy of the nineteenth century. This paper will examine a few of the principal aspects of the new economy of collective bargaining. It will aim to define and clarify issues rather than to suggest policies, but a few scattered comments on problems of policy will be included. Four principal topics will be discussed:

1. The effect of unions upon the administration of business enterprises.
2. The effect of wage determination through collective bargaining upon the operation of the economy—especially upon the volume of employment and output and the level of prices.
3. The problem of protecting the community from the interruption of vital services.
4. The problem of controlling the nature of the bargaining process.

II

How will collective bargaining affect the administration of business enterprises? Two basic issues arise. One is whether trade unions, possessing their own objective, their own policies, and their own leaders, will produce a rivalry for power and influence between unions and management which will prevent employees and management from working effectively as a team. The second question is how collective bargaining will affect the extent to which the administration of business concerns reflects the interests of consumers and the extent to which it reflects the interests of employees.

The answer to the first question depends in the main upon the nature of the bargaining process—a matter to be discussed separately. If collective bargaining turns out to be merely a process by which each side endeavors to assert its economic power, it will prevent the development of satisfactory teamwork between employees and management. The effect of collective bargaining upon teamwork also depends upon the success of managements and unions in reaching agreement concerning the appropriate functions of managements and unions in the operation of plants. One essential function of unions is to participate in policy-making—to help formulate the rules which guide the day-to-day operations of enterprises. This participation of unions shows itself in the main in the terms of trade agreements. The execution of policies is the essential responsibility of management, and management needs opportunity to exercise skill and discretion in applying policies to particular situations. Employees, however, need opportunity to challenge the fairness or wisdom of the day-to-day decisions of management. The grievance machinery established under trade agreements gives them an opportunity to do this.

Arrangements of the sort which I have indicated are common and, as a rule, have worked satisfactorily. Trouble arises if policies are formulated in excessive detail so that managements are deprived of reasonable opportunity to exercise their skill and their judgment. Trouble arises also when the execution of policies is made a joint responsibility of managements and unions. This means administration by bargaining and by joint committees instead of by management alone. It is cumbersome, it prevents administrators from properly exercising their skill, and it violates the principle that authority should be commensurate with responsibility—it prevents administrators who are responsible for output and costs from using their

best judgment in executing policies. This does not mean, of course, that administrators should not receive the help of advisory committees on which the unions may be represented. Trouble also arises when the employees are not given adequate opportunity to challenge the day-to-day decisions of management after these decisions have been made and applied. Management cannot expect to have discretion unless it is willing to permit its decisions to be reviewed in the light of specific facts and in the light of experience. Finally, trouble arises if the grievance machinery is abused and an effort is made by either side to obtain through administration conditions which could not be gained through negotiations.

Skilful management may cause unions to assist in the development of teamwork between employees and supervisors. Unions are often an excellent agency of communication—useful in helping managements learn more about the problems of employees and in helping employees learn more about problems of management. If management shows an interest in the problems of employees, there is a good chance that the employees will become interested in many problems of management, such as quality control and cost reduction. Likewise, there are some matters which are the joint concern of management and employees. Prevention of accidents and reducing the causes of labor turnover are examples. Working together on matters of mutual concern helps, of course, to develop the willingness and capacity of the management and the union to take an interest in each other's problems.

How will collective bargaining affect the extent to which the administration of enterprises reflects the interests of consumers and the interests of employees? The consumers' interest is in methods of production and administration which make possible low money costs; the interest of employees is in good working conditions, security of employment, a comfortable place of work, and the avoidance of too much competition among workers. The net product of industry is maximized when an appropriate balance is struck by the interests of consumers and the interests of employees. In this brief discussion I omit the interests of business owners which, though somewhat distinct, tend in labor matters to coincide in the main with the interests of consumers.

Until recently the administration of business concerns has undoubtedly been too favorable to consumers. Collective bargaining has tended in most situations to introduce a better compromise between the interests of consumers and the interests of workers. For example, it has compelled administrative policies to be based upon the more complete and more informed consideration of facts —particularly facts about working conditions. In addition, it has given employees the equivalent of civil rights within the plant—protecting them against arbitrary discharge or layoff. Most people would probably agree that these changes are for the better. The unanswered question is whether collective bargaining in the long-run will shift the balance too far in favor of the employees. Three principal questions arise:

1. Will collective bargaining enforce mechanical methods of selecting personnel so that managements will not be given reasonable opportunity to make use of the discoveries of physiologists and psychologists concerning the adaptability of different types of people for different types of work?

2. Will it substantially narrow opportunities for exceptional individuals in order to provide security for the many?

3. Will it seriously limit the total output of industry in order to make jobs in particular situations?

Only experience will give the answers to these questions. It is important, however, to see clearly that the problem exists. The more plainly it is seen, the better is the prospect that satisfactory compromises will be reached. In order to maintain a proper balance between the interests of employees and the interests of consumers, the community may have to determine whether certain provisions in trade agreements, such as those compelling wasteful utilization of labor, are contrary to public policy.

III

How will wage determination under collective bargaining affect the operation of the economy? This general problem breaks down into many parts of which three are particularly important:

1. The effect upon the rate of technological change.
2. The effect upon the structure of wages.
3. The effect upon the relationship between wages and prices, and consequently, the effect upon the level of employment and the movement of prices.

1. The pressure of trade unions for higher wages is likely to stimulate the growth of

industrial research. The more rapidly wages are expected to rise, the more enterprises can afford to spend today in order to use less labor tomorrow. Even a small tendency for collective bargaining to accelerate the rise in output per manhour can substantially raise the standard of living of the country. At the end of twenty years a 2 per cent annual increase in output per manhour means a total rise of 48.6 per cent, but a 3 per cent rate of increase means a total rise of 80.6 per cent, and a 4 per cent rise, 119.1 per cent.

2. Collective bargaining may be expected to have important effects upon the structure of wages. Much will depend upon the extent to which wage settlements reflect the relative bargaining power of employers and unions—that is, their relative willingness to endure a shutdown in order to enforce their terms. In free markets, workers of a given degree of skill and responsibility tend to receive the same real wages, with differentials which reflect variations in the attractiveness of jobs, occupations, or localities. Such a wage structure, indeed, is necessary to produce the most advantageous distribution of labor. Only by accident, however, would a wage structure which reflects the relative bargaining power of employers and unions bring about the optimum distribution of labor between industries, occupations, and places.

Is the maldistribution of labor which will occur if collective bargaining sets wages upon the basis of bargaining power likely to limit seriously the size of the national product? This is a crucial question on which research is badly needed. My guess is that the productivity of no more than one-third of the labor force would be substantially diminished by the maldistribution of labor and that the output of this one-third would not be reduced by as much as one-fourth, making a total loss of about 8.5 per cent of the national product. To this figure should be added some loss due to the creation of "wage distortion" unemployment—the kind of unemployment which occurs when too many people are attracted to certain industries, occupations, or places by wages which are abnormally high for the skill and responsibility involved and the conditions under which work is done. A wage structure which reflects the bargaining power of the parties will create wage distortion unemployment in those industries, occupations, or places where unions are particularly strong, but no one knows how large it is likely to be. It will depend upon the elasticity of the supply of labor in relation to the price of labor and this, in turn, will depend upon a number of conditions which have not been adequately explored.

3. How will collective bargaining affect the relationship between wages and prices? This is the most important and interesting question of all. It seems clear that collective bargaining will substantially change the relationship between wages and prices and that, as a result, the operation of the economy will be modified in important ways. It is not clear, however, just what the effect of collective bargaining upon wage-price relationships will be.

Prior to the rise of trade unions, the active determinant of wages during periods of expansion was the demand for labor. The supply price of labor has been a relatively passive influence. Hence wages have been a sluggish price, lagging behind commodity prices. During periods of expansion the lag of wages has helped investment opportunities to outrun the flow of investment-seeking funds. In consequence, most periods of expansion have been accompanied by a considerable growth of credit. One result has been the accentuation of the cyclical movements of business. As long as new debts were being incurred faster than old ones were being paid, the economy was stimulated. On the other hand, when contraction occurred, the drop in production and employment was made more severe because part of incomes were used to pay debts rather than to buy goods.

During periods of contraction the supply price of labor had a far more positive influence upon the price of labor. At any rate, the drop in the demand for labor produced little effect upon wages. To this fact, along with the great capacity of the economy to increase the money supply is to be attributed the failure of the technological revolution of the nineteenth century to produce a substantial drop in prices. Between 1840 and 1940 output per manhour of labor employed outside of agriculture rose roughly six-fold. The index of wholesale prices was 10 per cent higher in 1940 than in 1840, but since the index has an upward bias, the real level of prices was moderately lower in 1940 than in 1840. Hourly earnings of non-agricultural workers, however, were about eight times as high in 1940 as in 1840. In other words, virtually all of the gains of technological progress went to employees rather than to consumers. This happened during a period when only a small fraction of wage earners were organized and when the influence of unions upon wages was slight.

The rise of unions will cause the supply price of labor to become a much more active determinant of money wages. One possibility is that during periods of expansion the labor supply function will cease to be pretty largely independent of the demand function and that rightward shifts in the demand function will quite promptly produce leftward shifts in the supply function. As a result, the lag of wage increases behind price increases will be reduced. The experience of the last two years tends to support this conclusion. A second possibility is that the supply function will remain quite independent of the demand function but that changes in the supply price of labor will precede changes in the demand for labor. This would be a revolutionary change in the operation of the economy during periods of expansion. It would require that prices adjust themselves to wages rather than wages to prices—something which heretofore has happened only during periods of contraction. A third possibility is that there will be a mixture of the first two possibilities. In new industries where technological change is most rapid, shifts in the demand function will continue to precede shifts in the supply function, but the supply function will have lost its independence of the demand function. In more or less static industries, on the other hand, shifts in the supply function in a collective bargaining economy may be expected to occur in advance of rightward shifts in the labor demand function, and probably even in the face of no expectation of a rise in the demand for labor. Let us examine briefly, first, the possibility that collective bargaining will reduce the lag of wage increases behind price increases during periods of expansion and, second, the possibility that collective bargaining will cause increases in money wages to precede increases in prices.

If the lag of wages behind prices is reduced, important modification may be introduced into business cycles and into the susceptibility of the economy to cyclical movements. A reduction in the lag of wages behind prices will reduce both the volume of investment opportunities and the volume of investment-seeking funds. Unless reduction in the lag of wages behind prices accentuates the rise in prices, it will affect investment opportunities more than the volume of investment-seeking funds. Consequently, there is a strong probability that periods of expansion will be accompanied by a smaller expansion of credit than was characteristic of the economy before the growth of trade unions. In that event, the stability of the economy would be increased—though possibly at the cost of a lower average level of employment. This is a question which requires special investigation. There is reason to suspect that the unfavorable effects upon employment might be small or even non-existent. The rising prices which have accompanied many periods of expansion suggest that the demand for goods was growing faster than the capacity of the economy to increase employment. If this were so, a slower shift in demand functions might have only price effects and little or no employment effects.

A reduction in the lag of wages behind prices, however, could stimulate the expansion of credit. This might happen because of the effect of wage increases upon the timing of business expenditures. Wage advances over a large part of industry might increase the expectations of higher prices and cause business managers to step up their expenditures on inventories, plant, and equipment in order to buy before prices rose. If the volume of investment-seeking funds were reduced by the smaller lag of wages behind prices, the increase in business expenditures would have to be accomplished by the greater use of credit. In that event, the smaller lag of wages behind prices would accentuate rather than moderate the cyclical fluctuations of business.

Let us consider the possibility that trade unions will cause wage increases to precede price increases. Will prices adjust themselves to wages or will unemployment rise until unions are no longer able to raise labor costs relative to prices? Economists are not yet prepared to answer this question. During a considerable part of a period of expansion, prices may be highly responsive to increases in wages. For example, so long as price increases are expected for various reasons independent of wage changes, advances in wages are likely to reenforce these expectations and will stimulate buying in anticipation of higher prices. Difficulty may be encountered later, because buying in anticipation of needs cannot be expected to increase indefinitely. When buying for this reason falls off, a downward spiral may be started. Much will depend upon what has been happening to wages and prices. If prices, which by hypothesis started to rise *after* wages, increase sufficiently *faster* than wages, the increase in investment opportunities may so exceed the increase in investment-seeking

funds that even a drop of buying in anticipation of demand produces no drop in spending and in prices. Obviously, this result would be a happy accident.

If conditions which are independent of wage changes do not cause business men to expect prices to advance, the rise in wages may not be followed by an offsetting increase in prices. If the rise in wages fails to create the expectation of an offsetting rise in prices, there will be a negative shift in the investment function. Unless the immediate rise in wages relative to prices produces a sufficient negative shift in the saving function, expenditures will not increase sufficiently to cause prices to advance in proportion to the rise in wages or to maintain the previous volume of employment. In fact, expenditures may drop instead of increasing. The past behavior of the economy suggests that, when price increases are not expected for reasons independent of wage changes, the immediate effect of a rise in wages relative to prices is to produce a greater negative shift in the investment function than in the saving function. Were this not so, periods of expansion would not so readily run into periods of contraction or periods of contraction into periods of expansion.

The foregoing analysis indicates that the rise of trade unions will produce major changes in the operation of the economy, but that the precise nature of these changes cannot be predicted. One forecast may be ventured. If the upward pressure of wages upon prices should prove to be deflationary rather than inflationary, the government may be expected to stimulate a rise in prices. This leads one to conclude that in the collective bargaining economy of the twentieth century, the movement of prices will be more strongly upward than in the free market economy of the nineteenth century—either because the upward pressure of unions on wages under most circumstances is inflationary or because the government is forced to intervene in order to make prices rise as fast as wages. This is a most important statement. Suppose that prices need to rise only 2 per cent a year to offset the tendency of money wages under collective bargaining to rise faster than output per manhour. In ten years that would mean a 25 per cent advance; in forty years, an increase of more than 100 per cent. In other words, the life insurance or the pension which a man begins paying for at twenty-five years of age will have lost more than half its purchasing power by the time he reaches sixty-five. Collective bargaining is likely to require radical changes in the methods of financing life insurance and pensions.

IV

The rise of trade unions puts the community at the mercy of small groups which are able to cut the country off from vital services. So far these groups have shown considerable restraint in exercising their great power, although within the last two years the nation has four times been threatened with economic disaster because of a nation-wide shutdown of railroad transportation, coal mining, or steel making. How can the community protect itself against strikes or lockouts which would jeopardize its vital interests—its health, its safety, its ability to carry on production?

Some trade union leaders say that the right to strike is one of the basic institutions of free enterprise. No man should be compelled to work for the profit of another, they assert, on terms which are unacceptable to him. Consequently, depriving men of the right to strike is said to require that the government regulate profits. That means government regulation reaching to the very heart of industrial operations.

These assertions misconceive the nature of the right to strike. Prohibiting strikes does not mean that men are prevented from quitting or are compelled to work against their will. It simply means that they must choose between working and quitting—that a man is not permitted to hold a job at which he is unwilling to work because he does not like the wage or other conditions of employment. Furthermore, no government can recognize the "right" of small or large groups, whether they be bankers, manufacturers, or wage earners, to combine for the purpose of depriving the community of essential services. Certainly a strike or lockout is a trivial reason for jeopardizing the health and safety of the community, because it is merely an attempt to discover which of two sides has the greater economic strength. There is no reason to expect the strike or lockout to result in a just settlement. No less adequate reason for jeopardizing the health and safety of the community could be imagined than an attempt of the economically strong to impose their will upon the economically weak.

What is to be done? Merely concluding that the public health or public safety should not be jeopardized in order to permit two parties to test

their economic strength does not make railroad trains run or get coal mined. Nevertheless, the community must have protection against the stoppage of vital services.

Perhaps the problem will solve itself. No group likes to assume the responsibility for jeopardizing the welfare of the entire community. Certainly most union leaders are reluctant to order stoppages which would produce a national catastrophe. Consequently, union leaders may become interested in working out arrangements which will protect them from having to choose between acceptance of the employer's best offer and ordering a strike which would be a disaster for the country. An escape from this dilemma can be provided through arrangements which provide for arbitration—procedural agreements under which the parties agree to refer deadlocks to arbitration and which set up rules for selecting arbitrators and for conducting the proceedings.

If employers and unions in vital industries fail to provide arrangements for protecting the community, the government will undoubtedly have to act. Two or three more nation-wide shutdowns in vital industries would convince the community of this necessity. I do not propose to discuss here what the community might do, but the provisions of the Taft-Hartley Act on this point are unacceptable. These provisions authorize a government-conducted vote on the employer's "best offer" before the employees strike. It is obviously ridiculous for the government to conduct a strike vote among employees whose quitting would jeopardize the public health or the public safety.

V

What is to be the nature of the bargaining process under collective bargaining? This is the most fundamental question raised by the new economy of collective bargaining. Upon the answer to it will depend in large measure the effects of collective bargaining upon employer-union relations, upon the structure and behavior of wages, and hence the operation of the economy, and upon social relations within the community.

Is collective bargaining to be primarily a process of settling wages and conditions of employment on the basis of economic power? Or is it to be a process by which the parties seek to determine what is fair by some generally acceptable standards of fairness? Surprisingly little thought has been given to this problem. Early thinking about collective bargaining assumed that if employees were given the right to organize, rough equality of bargaining power would be created between business concerns and their employees and that reasonably fair bargains would result. Implicit in this thinking was the assumption that collective bargaining would fix the terms of the labor contract on the basis of economic power.

This is a dangerous kind of thinking. It is easy to show, of course, that bargaining on the basis of power might produce industrial peace. This result would follow provided neither side underestimated the other's willingness to fight. Bargaining of this type, however, could not be expected to produce cooperative relations between employers and unions. On the contrary, it would produce a sort of cold war or armed truce between employers and employees. It would give each side an incentive to impress the other with its toughness and its willingness to endure a shutdown in order to enforce its terms. Appeals to principle and the introduction of arguments and evidence would be more or less irrelevant. Parties dealing with each other on the basis of power could scarcely be expected to develop much interest in each other's problems or in the effect of their policies upon each other or upon the community. In short, the capacity of the community to deal unemotionally and realistically with many matters affecting business or labor would be impaired.

What will determine the nature of the bargaining process? This is a problem which the community at large must help to settle. It is not sufficient that each party to the bargain seek to impose on the other terms which the first party regards as fair. The ideas of the two parties may be quite far apart. There is needed a well-developed body of thought which represents more or less well-accepted views concerning what is fair. Consider, for example, the problem of determining the fairness of wage claims. An enterprise receives demands for a wage increase on the ground that its rates are below the prevailing rate in the community. Management rejects the demand on the ground that it is paying more than its highest-paying competitor. Of what relevance are these facts in determining the issue? Or consider the problem created by a rise in the cost of living—a common basis for wage demands. Advancing living costs mean that demand is outrunning supply. Would wage increases raise consumer purchasing power, aggravate the tendency

for demand to outrun supply, and thus cause living costs to rise even faster than ever?

Complete agreement between the employer and the union with respect to the criteria which should govern the settlement of wages in specific cases is not necessary. If negotiations, however, are to be an appeal to principle and an attempt to apply ideas of fairness to the specific situations, there must be a well-developed and more or less accepted body of thought upon which each side can base its arguments. An additional development of the needed body of thought may be necessary—namely, insistence by the community that any party which seeks to impose conditions on the other be willing to submit its claims to arbitration by neutrals. This need not be a legal requirement—at least in industries which do not provide vital services. Nevertheless, it is doubtful whether collective bargaining can be prevented from being a mere process of settling issues on the basis of power unless the community imposes strong moral pressure on each side to submit deadlocks over the terms to new agreements to arbitration.

VI

The foregoing analysis indicates that the central problem raised by the replacement of the free market economy of the nineteenth century by the collective bargaining economy of the twentieth century is how to get the interests which all members of the community have in common properly represented in the process of collective bargaining. Not since the late Middle Ages has the community been confronted with such highly organized small groups—groups which are necessarily motivated in part by parochial and particularistic interests and yet which profoundly affect the nature of business administration, the structure of wages, the relation of wages to prices and employment, and other matters of concern to the community as a whole.

Social control, of course, is not necessarily established by laws or public policies. In the main, it is accomplished through ethical codes, through conceptions of what is right or wrong, conceptions which are the heritage of all members of the community and which, in consequence, are accepted by them more or less as a matter of course. Hence the task of the community is one of developing important additions to its ethical code so that employers and trade unions will properly respect the interests which all members of the community have in common. To the extent that ethical codes fail effectively to get the interests of the community properly represented in collective bargaining, the country will need to control employers, unions, and collective bargaining by public policies, laws, and governmental institutions.